UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

Departamento de Física Atómica, Molecular y Nuclear

TESIS DOCTORAL

Irreversibility and dissipation in microscopic systems

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Édgar Roldán

Director

Juan M. R. Parrondo

Madrid, 2013

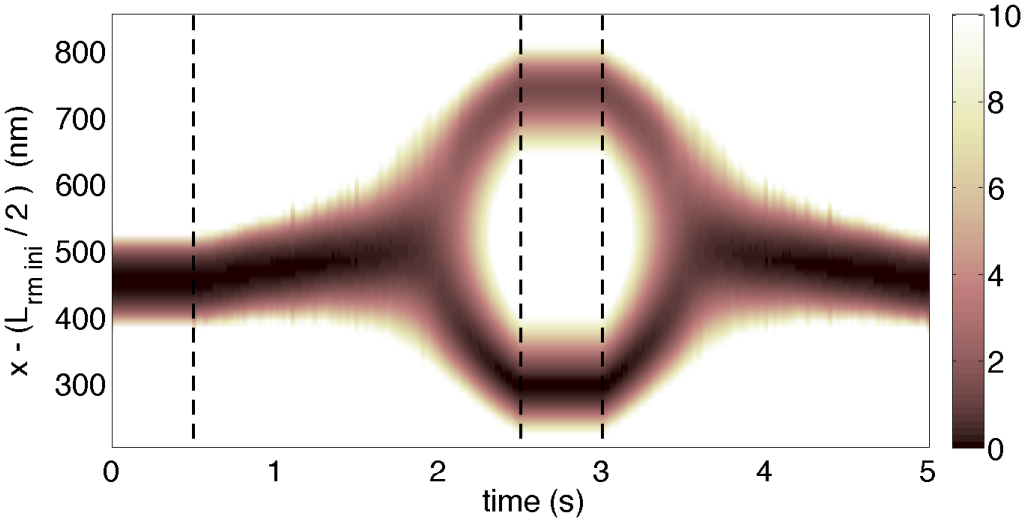
ÉDGAR ROLDÁN

IRREVERSIBILITY AND DISSIPATION IN MICROSCOPIC SYSTEMS

[May 28, 2013 at 1:09]

IRREVERSIBILITY AND DISSIPATION IN MICROSCOPIC SYSTEMS

ÉDGAR ROLDÁN



PhD thesis

Supervisor: J. M. R. Parrondo
Departamento de Física atómica, molecular y nuclear and GISC
Facultad de Ciencias Físicas
Universidad Complutense de Madrid

May 28, 2013, Version 1.0

ABSTRACT

The relationship between dissipation and irreversibility is at the core of statistical mechanics and thermodynamics [Cal85]. The first works on this relationship were done in linear irreversible thermodynamics, where the entropy produced by macroscopic systems weakly perturbed away from equilibrium was studied. Near equilibrium, the entropy production is proportional to the macroscopic currents or flows of the system such as the particle current or the heat flow [Kon98]. In this regime, time irreversibility is revealed as a presence of currents in the system. The connection between dissipation (or entropy production) and irreversibility (currents or flows) is merely qualitative in linear irreversible thermodynamics, and cannot be trivially generalized to microscopic systems driven arbitrarily far from equilibrium. In addition, entropy production and other thermodynamic quantities such as heat or work done by a microscopic system in nonequilibrium processes were not understood until the introduction of stochastic thermodynamics.

Recent works in the framework of nonequilibrium statistical mechanics have studied the extension of the laws of thermodynamics to small-scale (microscopic or smaller) systems, where thermal fluctuations play an important role in the dynamics. These new results are part of a new field in statistical mechanics called *stochastic thermodynamics*. Some of its most relevant results are the generalization of the first law to microscopic systems by defining the work and heat associated to a single stochastic trajectory or realization of a nonequilibrium process [Sek98]. The entropy of a microscopic system along a stochastic trajectory and the generalization of the second law to overdamped Brownian particles was done in [Sei05]. Both first and second law of thermodynamics are valid in the microscopic scale when ensemble averages of stochastic work, heat and entropy are considered.

Fluctuations in small systems give rise to rare events that do not happen in the macroscopic world, such as an entropy decrease in the universe or negative dissipation in a single realization. *Fluctuation theorems* are exact relationships, within the framework of stochastic thermodynamics, that express universal properties of the probability distribution of thermodynamic functionals of the stochastic trajectories of the system, such as work, heat or entropy [Sei12]. Crooks [Cro99] fluctuation theorem, Jarzynski's equality [Jar97] or the steady state fluctuation theorem [Eva93] are some of the most celebrated examples. Unlike linear irreversible thermodynamics, fluctuation theorems are

valid for processes where the system is driven arbitrarily away from equilibrium.

Recent advances in experimental techniques have allowed to measure physical magnitudes of small systems with nanometer and piconewton resolution in position and force, respectively, which amounts a precision of kT in the measurement of energy [Bus05]. In the majority of such experiments, optical tweezers, first described in [Ash86], are used to trap a microscopic particle and to drive it out of equilibrium. Optical tweezers have been used to experimentally test Jarzynski's equality [Lip02], Crooks fluctuation theorem [Col05] or steady state fluctuation theorems [Wan05a].

Exploiting the theoretical results derived in the context of fluctuation theorems is one of the main objectives of the current research in stochastic thermodynamics. In a recent work by Kawai *et. al* [Kaw07], the dissipation along a nonequilibrium processes for isolated microscopic systems that are initially connected to a thermal bath is investigated. For such processes, it is found that the average dissipation can be expressed in terms of a relationship between the phase space densities of the forward and the time-reversed (backward) process. Their formula, known as the *KPB theorem*, expresses the entropy production in terms of the microscopic properties of the system. The result also implies a connection between dissipation and the distinguishability between the forward and the backward process via a metric, the Kullback-Leibler divergence (KLD), that measures such distinguishability [Kaw07, Par09]. The KLD between the forward and backward phase space densities is a measure of the arrow of time. Consequently, this result is the first quantitative relationship between dissipation (or entropy production) and irreversibility (KLD between forward and backward phase space densities) derived for microscopic systems driven out of equilibrium.

The KPB theorem is found to be applicable to many different scenarios [Par09], and can be used to reproduce relevant results in the context of information thermodynamics. For example, Landauer's principle, which establishes the minimum amount of work needed to erase one bit of information, is a particular case of the KPB theorem. The energetics of the Szilard engine, which is a simplified one-particle Maxwell demon, can also be reproduced using the KPB theorem. An application of the result to microscopic systems that reach a nonequilibrium stationary state (NESS) was also suggested in [Par09]. A first experimental test of the relationship between dissipation and the KLD in the NESS was done by dragging Brownian particles with an optical tweezer in [And07].

In this thesis, we study the relationship between dissipation and irreversibility for microscopic systems that reach a NESS within the framework of stochastic thermodynamics. We are first interested in expressing this relationship in the NESS and secondly, we aim to apply this result to practical situations. The main goal of this work is to find a technique that allows one to estimate the average dissipation of a microscopic system in a NESS from the statistics of a single stationary trajectory produced by the system during the process. We want our technique to be able to estimate the average dissipation of the system by using any data (one or several discrete or continuous degrees of freedom) sampled from the system and even ignoring any physical detail of the system. Our method uses the KLD to measure the distinguishability between a stationary trajectory and its time-reversal, providing a quantitative tool to measure the arrow of time in the NESS, i.e. the time irreversibility.

Our work is a continuation of the theoretical result first proved in [Kaw07] and of the experimental test done in [And07]. We are interested in going one step further, by designing a tool to estimate the dissipation of any physical microscopic system that reaches a NESS, focusing on biological systems, where nonequilibrium processes play a fundamental role. We want our technique to be able to capture the irreversibility of a process even when having access only to partial information of the system (*i.e.* when only a subset of the phase space can be sampled), which is the most common situation in the experimental context. In addition, we aim to explore the relationship between irreversibility and dissipation in the NESS by designing new experimental protocols using optical tweezers.

We first clarify in Chapter 2 the relationship between dissipation and irreversibility for microscopic systems that reach a NESS, by expressing the average entropy production in a NESS in terms of the KLD rate between the probability to observe a trajectory and the probability to observe the time-reversed trajectory. We show that this KLD rate can only be calculated analytically for some specific processes such as a Markov chain obeying detailed balance or Hidden Markov chains. For general stochastic processes, the KLD rate has to be estimated from empirical data.

In order to apply our technique to a variety of physical situations, we develop in Chapter 3 estimators of the KLD rate for both discrete and continuous data. We find techniques that allow estimating the KLD using finite data from a single stationary trajectory. For discrete time series, using statistics of blocks or sequences of $m-$data is found to be the most feasible technique to estimate the KLD rate as long as there is sufficient statistics, as described in Chapter 3.1.1. In Chapter 3.1.2, we introduce a second estimator for discrete data based on zippers or compres-

sion algorithms that outperforms the estimator based on block entropies when there is not sufficient statistics. Additionally, we introduce two new different techniques to measure the KLD rate for real-valued or continuous time series. First, a novel estimator that transforms a time series into a directed graph whose in-going and out-going degree distributions capture the irreversibility of a time series (see Chapter 3.3). This method distinguishes between reversible and irreversible generic stochastic processes as well as between conservative and dissipative chaotic processes. We also introduce in Chapter 3.2.2 a second estimator of the KLD rate for continuous correlated time series which uses an asymmetric functional to produce a new uncorrelated series whose KLD is easier to calculate. This estimator can be easily connected to the average dissipation of a microscopic system in the NESS, which is analyzed using biological data in Chapter 5.

We first apply our technique to a discrete flashing ratchet model (see Chapter 4), analyzing the bound to the dissipation provided by the KLD rate when having access to trajectories containing full or partial information of the physical system. We show that, even when having access only to partial information of the system, we can distinguish between equilibrium and NESS by means of our KLD rate estimator. When an external force stalls the ratchet, no net current of particles is observed and linear irreversible thermodynamics fail to predict that the system is dissipating energy. However, using out technique, we correctly predict that the system is producing entropy despite no net current is observed. In this simple example, the KLD, and not the current, is revealed as the actual fingerprint of irreversibility in microscopic systems.

Our next step is to apply our technique to real data obtained from a biological system that produces a real-values stationary time series. We aim not only to distinguish between equilibrium and NESS, but also to distinguish between two different types of NESS that are present in biology: those produced by active processes and those that are passive. In Chapter 5 we apply our techniques to experimental data of bullfrog's ear hair bundles. Our KLD rate estimator based on the use of residuals distinguishes between active and passive spontaneous oscillations of ear hair bundles from bullfrog's sacculus. Our technique only requires the measurement of spontaneous oscillations of the position of the top of the hair bundles during $\sim$ 100s, outperforming significantly the requirements of previous works [Mar01b]. In addition, our KLD rate estimator indicates the minimum irreversibility to sustain spontaneous oscillations in bullfrog's ear hair bundles, which is of the order of the energy consumption rate from ATP hydrolyzation by a single adaptation motor.

The last part of the thesis is devoted to the theory and experiments of other interesting phenomena of stochastic thermodynamics. The energetics of a symmetry breaking and a symmetry restoration can be linked to the relationship between dissipation and irreversibility in the microscopic scale. When a microscopic system breaks a symmetry, the accessible phase space of the system is reduced to a subset of the full phase space. Even when the symmetry is broken quasi statically, the symmetry breaking is accompanied by a (negative) entropy production. We have derived a formula that relates the average entropy production when the system chooses an instance with the probability of the system to choose that option. Our formula is valid in the quasi static limit and it is universal: it does not depend on the physical nature of the system or on the mechanism that biases the probability to choose any option. In addition, we have obtained an equivalent formula for the symmetry restoration process, which, together with the symmetry breaking result, can be used to derive Landauer's principle and the energetics of the Szilard engine as discussed in Chapter 6. For a complete validation of our formula, we have checked our results experimentally, by trapping a Brownian particle in a dual optical tweezer and using external electric fields to tune the probability of the bead to choose between the two traps.

Finally, as another experimental application of stochastic thermodynamics, we have designed an experiment to tune the amplitude of the Brownian fluctuations of a microscopic charged sphere. The setup consists of a Brownian particle trapped by an optical tweezer and an external random electric field of tunable amplitude. With our setup, we are able to control the kinetic temperature of a microscopic particle from room temperature to 3000K in a fast timescale ($\sim$ ms). This technique opens the possibility of implementing heat engines at the microscopic scale such as a microscopic non-quasistatic Carnot engine.

To conclude, in this thesis, we have shown that the time reversal asymmetry of a stationary time series provides information about the entropy production of the physical mechanism generating the series, even if one ignores any detail of that mechanism. We have performed a complete study on the relationship between dissipation and irreversibility for microscopic system in the NESS. Our study includes analytical results, simulations and experimental verifications. Our results provide new insights in stochastic thermodynamics and nonequilibrium statistical mechanics but also in time series analysis, information theory and biophysics.

Our work has potential applications from an experimental point of view. Using our technique, one can, from the data of a single (discrete or continuous) stationary trajectory obtained sampling any physical variable of a microscopic system, estimate

the average dissipation rate of the mechanism that generated the data. This opens the possibility to distinguish between different nonequilibrium behaviors from biological data using the data from only one experimental sampling.

On the other hand, our experimental results using optical tweezers have provided a new insight in the thermodynamics of choice, finding exact relationships between entropy production and the probability to choose an option in a symmetry breaking. This could be used to estimate the probability of a biomolecule to adopt a specific conformation from energetic measurements. On the other hand, the implementation of adiabatic processes in the microscopic scale — and therefore the construction of a Carnot engine — is also possible by using active feedback when tuning the kinetic temperature of a Brownian particle with our experimental setup described in Chapter 7.

RESUMEN

La relación entre irreversibilidad y disipación está en la base de la mecánica estadística y la termodinámica [Cal85]. Los primeros trabajos sobre esta relación se realizaron en termodinámica lineal de procesos irreversibles, en los que se estudió la entropía producida por un sistemas macroscópicos débilmente perturbado fuera del equilibrio. Cerca del equilibrio, la producción de entropía es proporcional a las corrientes macroscópicas o flujos del sistema, tales como la corriente de partículas o el flujo de calor [Kon98]. En este régimen, la irreversibilidad temporal se revela como una presencia de corrientes en el sistema. La conexión entre la disipación (o producción de entropía) y la irreversibilidad (corrientes o flujos) es meramente cualitativa en termodinámica lineal irreversible, y no puede ser trivialmente generalizada a sistemas microscópicos perturbados arbitrariamente lejos del equilibrio. Además, la producción de entropía y otras magnitudes termodinámicas como el calor o el trabajo realizado por un sistema microscópico en procesos de no equilibrio no se fueron estudiados hasta la introducción de la termodinámica estocástica.

Trabajos recientes en el marco de la mecánica estadística de no equilibrio han estudiado la extensión de las leyes de la termodinámica a sistemas pequeños (microscópicos o de dimensión menor), donde las fluctuaciones térmicas juegan un papel importante en la dinámica. Estos nuevos resultados son parte de un nuevo campo en mecánica estadística llamado *termodinámica estocástica*. Algunos de los resultados mas relevantes en este campo son la generalización de la primera y la segunda ley de la termodinámica a sistemas microscópicos, definiendo el calor y el trabajo asociados a una trayectoria estocástica o realización de un proceso

de no equilibrio [Sek98]. En [Sei05], se estudió la entropía de un sistema macroscópico a lo largo de una trayectoria estacionaria y la generalización de la segunda ley a partículas brownianas sobreamortiguadas. Tanto la primera como la segunda ley de la termodinámica se recuperan en escala microscópica cuando se consideran promedios a realizaciones de trabajo, calor y entropía.

Las fluctuaciones térmicas dan lugar a fenómenos extraños en sistemas pequeños que no ocurren en el mundo macroscópico. En el mundo microscópico, es posible observar un decrecimiento de la entropía del universo o un valor negativo de la disipación en una trayectoria estocástica producida en un proceso de no equilibrio. Los *teoremas de fluctuación* son relaciones exactas que, en el marco de la termodinámica estocástica, expresan propiedades universales de la distribución de probabilidad de funcionales termodinámicos de la trayectoria estocástica del sistema, como el trabajo, el calor o la entropía [Sei12]. El teorema de fluctuación de Crooks [Cro99], la igualdad de Jarzynski [Jar97] o el teorema de fluctuación de estado estacionario [Eva93] son algunos de los ejemplos más destacados. Al contrario que en el caso de la termodinámica lineal irreversible, los teoremas de fluctuación son válidos para procesos en los que el sistema es perturbado arbitrariamente fuera del equilibrio.

Avances recientes en las técnicas experimentales han permitido medir propiedades físicas de los sistemas pequeños con precisión de nanómetros y piconewtons en posición y fuerza respectivamente, lo que implica una precisión de kT en la energía [Bus05]. En la mayoría de estos experimentos, las pinzas ópticas, descritas por primera vez en [Ash86], se utilizan para atrapar una partícula microscópica y sacarla del equilibrio. Las pinzas ópticas se han utilizado para verificar experimentalmente numerosos resultados de la termodinámica estocástica como la igualdad de Jarzynski [Lip02], el teorema de fluctuación de Crooks [Col05] o los teoremas de fluctuación de estado estacionario [Wan05a].

Uno de los principales objetivos de la investigación actual en termodinámica estocástica es la explotación y aplicación de los resultados teóricos derivados en los últimos años. En un trabajo reciente de Kawai *et. al* [Kaw07], se estudió la energía disipada en procesos de no equilibrio donde sistemas microscópicos aislados están inicialmente conectados a un baño térmico. Para esos procesos, se encontró que la disipación media se puede expresar en función de una relación entre la densidad del espacio de fases del proceso (hacia adelante) y el proceso bajo inversión temporal (hacia atrás). Su fórmula, conocida como el *teorema KPB*, expresa la producción de entropía en términos de las propiedades microscópicas del sistema. El resultado también implica una conexión entre disipación y distinguibilidad entre el proceso hacia adelante y el proceso hacia atrás a través de una

métrica, la divergencia de Kullback-Leibler (KLD), que mide esa distinguibilidad [Kaw07, Par09]. La KLD entre las densidades del espacio de fases del proceso hacia adelante y el proceso hacia atrás es una medida de la flecha temporal del proceso. Consecuentemente, este resultado es la primera relación cuantitativa entre disipación (o entropía producida) e irreversibilidad (KLD entre las densidades del espacio de fase del proceso hacia adelante y el proceso hacia atrás) obtenida para procesos de no equilibrio en sistemas microscópicos.

El teorema KPB es aplicable a muchos procesos de diferente naturaleza [Par09] y puede usarse para reproducir resultados relevantes en el marco de la termodinámica de la información. Por ejemplo, el principio de Landauer, que establece el trabajo mínimo para borrar un bit, es un caso particular del teorema KPB. La energética del motor de Szilard, una versión simplificada del demonio de Maxwell formado por una partícula, puede ser reproducida también por el teorema KPB. La aplicación del teorema KPB a sistemas microscópicos que alcanzan un estado estacionario de no equilibrio (NESS) fue sugerida en [Par09] y testada experimentalmente por primera vez desplazando partículas brownianas con una trampa óptica en [And07].

En esta tesis, estudiamos la relación entre disipación e irreversibilidad en sistemas microscópicos que alcanzan un NESS en el marco de la termodinámica estocástica. En primer lugar buscamos expresar esta relación en el NESS, y en segundo lugar buscamos aplicar este resultado a situaciones prácticas. El objetivo principal de este trabajo es encontrar una técnica que permita estimar la disipación media de un sistema microscópico en un NESS a partir de la estadística de una única trayectoria estacionaria producida por el sistema durante el proceso. Queremos que nuestra técnica sea capaz de estimar la disipación media del sistema a partir de cualquier serie de datos (de una o varias variables continuas o discretas) sampleadas del sistema y ignorando los detalles físicos del propio sistema. Nuestro método usa la KLD para medir la distinguibilidad entre la trayectoria estacionaria y su inversión temporal, y es una herramienta cuantitativa para medir la flecha del tiempo en el NESS, es decir, la irreversibilidad temporal.

Nuestro trabajo es una continuación del resultado teórico demostrado en [Kaw07] y del test experimental hecho en [And07]. Estamos interesados en ir un paso más allá, diseñando una herramienta para estimar la disipación de cualquier sistema microscópico que alcanza un NESS, haciendo hincapié en los sistemas biológicos, donde los procesos de no equilibrio juegan un papel fundamental. Nuestro objetivo es que nuestra técnica sea capaz de capturar la irreversibilidad de un proceso incluso cuando sólo se tiene acceso a información parcial del sistema

(es decir, cuando solo se puede samplear un subconjunto del espacio de fases), que es la situación más común en el contexto experimental. Adicionalmente, buscamos explorar la relación entre disipación e irreversibilidad en el NESS diseñando nuevos protocolos experimentales utilizando pinzas ópticas.

En primer lugar, clarificamos la relación entre irreversibilidad y disipación en el NESS para sistemas microscópicos en el NESS en el Cap. 2, expresando la entropía producida media en un NESS en términos de la KLD entre la probabilidad de observar una trayectoria y la probabilidad de observar la trayectoria invertida temporalmente. Demostramos que esa KLD solo se puede calcular analíticamente para casos específicos, como cadenas de Markov que obedecen balance detallado local o cadenas de Markov ocultas. Para procesos estocásticos genéricos, la KLD tiene que estimarse a partir de datos empíricos.

Para aplicar nuestra técnica a diferentes situaciones físicas, en el Cap. 3 introducimos nuevos estimadores de la KLD para trayectorias discretas y continuas. Hemos desarrollado técnicas que permiten estimar la KLD utilizando los datos de una única trayectoria estacionaria que contiene un numero finito de datos. Para series discretas, la técnica más efectiva es utilizar estadística de bloques de m datos, como se describe en el Cap. 3.1.1. En el Cap. 3.1.2, introducimos un nuevo estimador basado en zipping o la compresión de la trayectoria. Este método mejora la estimación de la KLD usando entropías de bloques de datos cuando no hay estadística suficiente en la serie temporal. Adicionalmente, introducimos dos nuevas técnicas para medir la KLD de series temporales que pueden tomar cualquier valor real (o continuas). En primer lugar, introducimos un nuevo estimador que transforma una serie temporal en un grafo dirigido, en cuyas distribuciones de grado salientes y entrantes se encuentra la información acerca de la irreversibilidad de la serie original (ver Cap. 3.3). Usando esta estimación de la KLD podemos distinguir entre procesos estocásticos reversibles e irreversibles y entre caos disipativo y conservativo. Además, en el Cap. 3.2.2 introducimos un segundo estimador de la KLD de series continuas correlacionadas temporalmente que usa un funcional asimétrico para producir una nueve serie descorrelacionada cuya KLD es más fácil de calcular. La relación entre este último estimador y la disipación es más clara que en el otro método, como se demuestra en el Cap. 5 usando series temporales de sistemas biológicos.

En primer lugar, aplicamos nuestra técnica a una flashing ratchet discreta (ver Cap. 4), analizando la cota a la disipación que da la KLD cuando se tiene acceso a trayectorias descritas por información total y parcial de la ratchet. Demostramos que, incluso teniendo sólo acceso a información parcial del sistema, es posible distinguir entre equilibrio y NESS a partir del valor

de nuestra estimación de la KLD. Cuando una fuerza externa para la ratchet, no se observa una corriente de partículas y la termodinámica lineal irreversible predice incorrectamente que el sistema está en equilibrio al no haber flujos. Sin embargo, usando nuestra técnica, predecimos correctamente que el sistema está produciendo entropía pese a que no se observa corriente de partículas. En este ejemplo simple, la KLD, y no la corriente, se muestra como la huella dactilar de la irreversibilidad en sistemas microscópicos.

Nuestro siguiente paso consiste en aplicar la técnica a datos experimentales obtenidos de un sistema biológico que produce series estacionarias. Buscamos no solo distinguir entre equilibrio y NESS, sino también distinguir entre diferentes tipos de NESS en biología, más concretamente, entre procesos activos y pasivos. En el Cap. 5, aplicamos nuestra técnica a datos experimentales de los hair bundles del oído de la rana. Nuestro estimador de la KLD basado en el uso de residuos, distingue entre oscilaciones espontáneas activas y pasivas en hair bundles del sáculo del oído de la rana. Nuestra técnica solo requiere medir oscilaciones espontáneas de la posición del hair bundle durante $\sim$ 100s, reduciendo drásticamente los requisitos necesarios de otros trabajos [Maro1b]. Además, nuestro estimador de la KLD estima la mínima irreversibilidad necesaria para sustentar las oscilaciones espontáneas, que es del orden de la energía consumida en hidrolización de ATP por unidad de tiempo.

La última parte de la tesis está dedicada a la teoría y experimentos de otros fenómenos interesantes de la termodinámica estocástica. La energética de una ruptura de simetría y de un restablecimiento de una simetría pueden obtenerse como un caso particular de la relación entre disipación e irreversibilidad para sistemas microscópicos. Cuando un sistema microscópico rompe una simetría, el espacio de fases accesible se reduce a un subespacio del espacio de fases total. Incluso cuando la simetría se rompe cuasi estáticamente, la ruptura de simetría está acompañada de una producción de entropía negativa. Hemos derivado una fórmula que relaciona el valor medio de la entropía producida cuando el sistema escoge una opción con la probabilidad de elegir esa opción. Nuestra fórmula es válida en el límite cuasi estático y es universal: no depende de la naturaleza física del sistema o del mecanismo que favorece una de las opciones. Además hemos obtenido una formula equivalente para el restablecimiento de una simetría, que, junto con la fórmula de la ruptura de simetría, puede utilizarse para reproducir el principio de Landauer y la energética del motor de Szilard, como se prueba en el Cap. 6. Para una validación completa de la fórmula, hemos comprobado nuestros resultados experimentalmente, atrapando una partícula browniana con dos pinzas ópticas y utilizando un campo eléc-

trico externo para modular la probabilidad de la partícula para escoger entre las dos trampas.

Por último, hemos desarrollado otra aplicación de la termodinámica estocástica, diseñando un experimento que es capaz de modular la amplitud de las fluctuaciones brownianas de una esfera microscópica cargada. El setup consta de una partícula browniana atrapada con una pinza óptica y un campo eléctrico externo aleatorio. Con nuestro setup, somos capaces de controlar la temperatura cinética de una partícula microscópica desde temperatura ambiente hasta 3000K en una escala temporal de ~ ms. Esta técnica abre la posibilidad de implementar motores térmicos de escala microscópica como un motor microscópico de Carnot.

En conclusión, en esta tesis hemos demostrado que la asimetría temporal de una serie temporal estacionaria da información acerca de la entropía producida por el mecanismo físico que generó la serie, incluso si no se conoce ningún detalle sobre dicho mecanismo. Hemos hecho un estudio completo de la relación entre disipación e irreversibilidad para sistemas microscópicos que alcanzan un NESS. Nuestro estudio incluye resultados analíticos, simulaciones y verificaciones experimentales. Nuestro trabajo aporta nuevas contribuciones en termodinámica estocástica, física estadística de no equilibrio, pero también en análisis de series temporales, teoría de la información y biofísica.

Nuestro trabajo tiene aplicaciones potenciales desde un punto de vista experimental. Utilizando nuestra técnica es posible, utilizando los datos de una única trayectoria estacionaria obtenida sampleando cualquier variable de un sistema microscópico, estimar la disipación del mecanismo que generó la trayectoria. Esto abre la posibilidad de distinguir entre diferentes comportamientos de no equilibrio en biología a partir de una única medida experimental.

Por otro lado, nuestros resultados experimentales utilizando pinzas ópticas sacan a la luz nuevos descubrimientos en la termodinámica de la elección, obteniéndose relaciones exactas entre producción de entropía y la probabilidad de elegir una determinada opción. Nuestra fórmula puede utilizarse para estimar la probabilidad de una biomolécula a adoptar una determinada conformación a partir de medidas energéticas. Por otro lado, nuestro dispositivo experimental usado en el Cap. 7 puede utilizarse para implementar procesos adiabáticos en escala microscópicas, y por tanto para construir un micro-motor de Carnot, utilizando feedback en el control de temperatura cinética.

PUBLICATIONS

Some ideas and figures have appeared previously or will appear in the following publications:

1. *Estimating dissipation with single stationary trajectories.* É. Roldán and J.M.R. Parrondo, Phys. Rev. Lett. **105** 150607 (2010).

2. *Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems.* É. Roldán and J. M. R. Parrondo, Phys. Rev. E **85** 031129 (2012).

3. *Time series irreversibility: a visibility graph approach.* L. Lacasa, Á. Núnez, É. Roldán, J.M.R. Parrondo and B. Luque. Eur. Phys. Journal B **85**, 6, pp. 1-11 (2012).

4. *Effective heating to several thousand kelvin of an optically trapped sphere in a liquid*. I. A. Martínez, É. Roldán, J.M.R. Parrondo and Dmitri Petrov. Phys. Rev. E **87** 032159 (2013)

5. *Universal features in the energetics of symmetry breaking*. É. Roldán, I. A. Martínez, J.M.R. Parrondo and Dmitri Petrov. In preparation.

6. *Quantifying irreversibility in biological time series as a tool to estimate dissipation: the case of oscillations in hair cell bundles*. É. Roldán, F. Jülicher, P. Martin and J. M. R. Parrondo. In preparation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

Part I

INTRODUCTION

1 SMALL-SCALE THERMODYNAMICS

The laws of thermodynamics apply to macroscopic systems, which are described by a large number of degrees of freedom ($N \sim 10^{23}$). Since the relative deviations from the mean behavior scale like $1/\sqrt{N}$, the deviations from the laws of thermodynamics are extremely unlikely to occur in the macroscopic scale [Jar11]. In the last decade, it has been possible to manipulate and measure physical systems at the micro and nano scale, such as colloidal particles [Duc91], DNA [Bus03] or molecular motors [Vis99]. At this scale, deviations from the mean behavior are more apparent. Moreover, the random forces exerted by thermal fluctuations are of the same order of magnitude of the forces that drive microscopic systems out of equilibrium. Notice that the order of magnitude of forces exerted by microscopic systems are of the order of piconewtons and the displacements of the order of nanometers, and the average energy of thermal fluctuations at room temperature is of the order of pNnm, $kT \simeq 4$pNnm for $T = 25^{o}C = 298.15$K.

The need to describe the dynamics of small systems has led to the introduction of *stochastic thermodynamics* [Sei08], which aims to extend thermodynamics to the regime where thermal fluctuations are relevant, that is, systems of small size (characteristic size of $\sim \mu$m) which undergo processes that can be sampled at submilisecond time resolution. By defining the stochastic heat and work associated to an stochastic trajectory [Sek10] the first law can be extended to microscopic systems driven out of equilibrium. The second law of thermodynamics presents deviations at this scale: the total entropy production can decrease in a single realization of the process. However, the average entropy production over many realizations of the process is positive, being the second law true only in average [Jar08].

In the last decade, there has been an increasing interest in describing laws that quantify the probability to observe deviations from the mean of thermodynamic quantities in nonequilibrium processes of microscopic systems. The *fluctuation theorems* provide quantitative relations that measure the probability to observe rare events concerning the work performed by the system [Jar97, Cro99] or the entropy production [Gal95]. In addition to that, fluctuation theorems open new possibilities in the microscopic scales, such as providing tools to measure the time irreversibility in a quantitative way or designing experiments that yield new fundamental results, as it is done in this work.

This chapter reviews the basic concepts used along the rest of the work. The chapter is organized as follows: In Sec. 1.1, we review the basic concepts of macroscopic thermodynamics, focusing on the first and second law of thermodynamics. In Sec. 1.3 we present some the main results of statistical mechanics that connect the thermodynamics of macroscopic systems with the statistics of their microscopic constituents. In Sec. 1.4, we review the fundamental aspects of stochastic thermodynamics that concern the energetics of microscopic particles, namely, the definition of stochastic work and heat and the expression of the first and second laws in the realm of stochastic processes. In Sec. 1.5 we discuss the theoretical framework of the fluctuation theorems connected to this work, and in Sec. 1.6 we show some of the most important experiments done in the recent years in stochastic thermodynamics.

1.1 REVIEW OF THERMODYNAMICS

Thermodynamics was originally developed as a part of physics that studies macroscopic systems, which are described by a large number of degrees of freedom, $N \sim 10^{23}$. In thermodynamic, the measurement of physical properties are restricted in both time and space [Cal85]: the measurements in the macroscopic scale are significantly slower than the atomic scale of time and significantly coarser than the atomic scale of distance. Examples of macroscopic systems described by thermodynamics are a gas in a piston, a streched wire or a paramagnetic rod [Fer56, Zem68, Cal85, Kon98]. Examples of properties of the systems are volume V, temperature T, pressure p or magnetization M. Macroscopic systems can exchange matter and energy with their surroundings. Depending on the boundaries between the system and the surroundings and their interaction, physical systems can be classified in three different types: [Kon98].

- **Isolated systems**: No energy and no mass exchange with the surroundings.
- **Closed systems**: They can exchange energy but not mass with the surroundings.
- **Open systems**: Both energy and mass can be exchanged with the surroundings.

Throughout this work, we will focus on open and closed systems, which are ones that are more present in nature.

The definition of system itself or its surroundings depends on the definition of the walls or boundaries that separate the system from the environment [Zem68], which can be *diatermic* (heat can flow through the wall) or *adiabatic* (heat cannot flow through

the wall). Within the surroundings, we can distinguish between two parts (see Fig. 2): The *thermal environment* (thermal reservoir, thermal bath) is defined as the physical system that surrounds the system and it is formed by a number of particles large enough to maintain its own temperature constant irrespective of the amount of energy exchanged with the system [Pat88]. The *external agent* is an external system that controls at least one of the degrees of freedom of the system. A simple example is a gas contained in a movable piston with diathermic walls. In this case, the gas is the system, the air surrounding the piston is the thermal environment and the experimentalist that controls the position of the piston is the external agent.

The state of the system at any time can be described by physical variables that are not independent, since they are connected by an *equation of state* of the form, *e.g.* for the case of a gas in a piston $f(p, V, T) = 0$. A particular case of state of a system are *equilibrium states*, in which the macroscopic variables of the system do not change and no macroscopic flows of energy and matter are observed.

When one or more of the macroscopic variables that describe a physical system change in time we say that a *thermodynamic process* occurs. In general, the evolution the evolution of a physical system is controlled by the external agent and modeled by the evolution of a *control parameter* $\lambda(t)$. According to the velocity at which the control parameter changes, the physical processes can be

- **Reversible processes**: At every time, the system is infinitesimaly near an equilibrium state. The control parameter is changed quasistatically (infinitely slow) and there is no energy dissipation.
- **Irreversible processes**: The control parameter is changed at finite rate and the system is not at every time in an equilibrium state during its evolution. Because of this, energy is dissipated to the thermal environment.

In practice, all the processes in nature are irreversible since it is experimentally impossible to change a control parameter quasistatically. However, one can use quasistatic processes as a model to describe the limit when the variation of the control parameter is slow. In Fig. 1 we show an example of a physical process that can be performed in both reversible and irreversible manner. An ideal gas is confined in a diathermic container that is in contact with a thermal bath at temperature T. The gas is initially in equilibrium at pressure p_A and volume V_A and its equation of state is the law of ideal gases $p_A V_A = NkT$, where N is the number of molecules of the gas. If the gas is compressed reversibly or infinitely slowly, the gas is in equilibrium along the whole process and its pressure

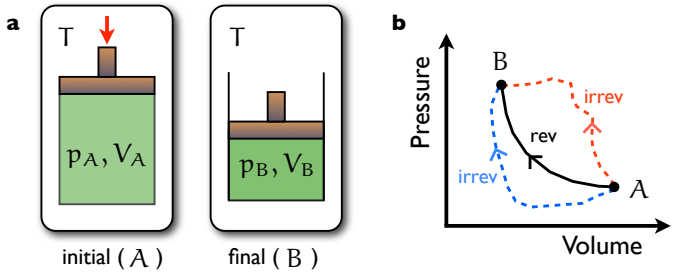


Figure 1: An example of an isothermal process of a closed system, the compression of an ideal gas confined in a container with diathermic walls. The surroundings are at constant temperature T and the container has a movable piston that can be held fixed at any position. **a**. Scheme of the process. The gas is initially in an equilibrium state A with initial volume V_A and initial pressure p_A. In the end of the process, the volume of the gas has been reduced to $V_B < V_A$ and its pressure has been increased to $p_B > p_A$. **b**. Representation of the physical process in the pressure-volume diagram. The black curve represents a reversible transformation during which the system is in an equilibrium state in which the gas satisfies $pV = NkT$, being N the number of molecules of the gas and k Boltzmann's constant. Blue and red dotted lines represent arbitrary irreversible processes than can be done between equilibrium states A and B.

and volume satisfy the law of ideal gases through the process. If the gas is compresses irreversibly, the gas no longer reaches an equilibrium state throughout the process and the law of ideal gases is not satisfied at any time between the beginning and the end of the process. Notice that there are many different possible ways to drive the system out of equilibrium whereas there is only one possible way to do the process reversibly.

On the other hand, according to the nature of the surroundings and the conditions of the experiment, the processes can be classified as follows:

- **Isothermal processes**: The system is in contact with the environment via diathermic walls and the temperature remains constant during the process.
- **Adiabatic processes**: The system is isolated from the environment and there is no energy (heat) exchange between the system and the surroundings.
- **Isochoric and isobaric processes**: The process is done at constant volume or pressure, respectively.

During a process, the total energy of a system changes in general. The *internal energy*, U of a system is defined as the

total energy of a system and it is the sum of the kinetic and the potential energy of all the atoms and electrons that form the system. During a process, part of the change of the internal energy of the system is due to the change of the control parameter by the external agent energy and the rest is exchanged with the thermal reservoir. The part of the energy that is transferred to the system that is due to the action of the external agent is called *work*, W, and the energy transferred from the thermal environment to the system is called *heat*, Q. The sum of these two contributions equals to the change in internal energy $\Delta U = U_{fin} - U_{ini}$,

First Law of Thermodynamics

$$\Delta U = Q + W. \tag{1.1}$$

The above relation is known as the *First Law Of Thermodynamics* Fig. 2 we sketch the first law of thermodynamics. The arrows

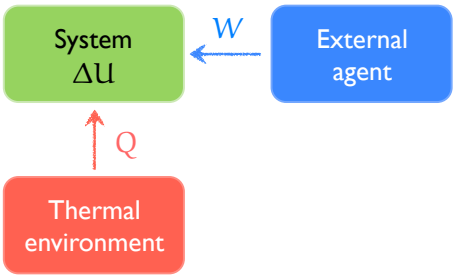


Figure 2: Illustration of the first law of thermodynamics. The heat Q is defined positive when it is transferred from the environment to the system and the work W is defined positive when it is done from the external agent to the system. The change in internal energy in the system is equal to the sum of the heat transfer plus the work done along the process, $\Delta U = Q + W$.

indicate the direction of the energy flow that corresponds with a positive sign of the heat and the work. When, for a closed system, an infinitesimal change of the control parameter is done, the internal energy of the system changes a small amount dU, and the first law applies as follows:

$$dU = \delta Q + \delta W \tag{1.2}$$

being δQ and δW the amount of heat and work that is transferred to the system when a differential change in the control parameter is done. Notice that δQ and δW are *inexact differentials*, which do not satisfy the usual properties of a differential. For a process in which a system starts in equilibrium state A and ends in another equilibrium state B, the integral of δQ along the process is the total heat transferred to the system $Q = \int_A^B \delta Q$. If the process is not reversible, this integral cannot be expressed as the value of a function (primitive) in B minus the value of the same function in A, since the heat Q depends not only on the

initial and final states but also on the path that the system follows in the coordinate space. This is the same that happens with the work and opposite to what happens with internal energy, where $\int_A^B dU = U_B - U_A$. For a cyclic process the change of internal energy is 0 since the initial and final states are the same irrespective of how the process is done, but not the heat and the work, which are trajectory dependent. In this case, by virtue of the first law $Q_{cycle} = -W_{cycle}$.

Although all the processes compatible with the first law are in principle possible to occur, not all of them are seen in nature. A glass falling to the ground smashing into pieces is an example. This process is likely to be observed in nature, but not its time reversal, the pieces going to its original position. The second law of thermodynamics expresses a restriction of some processes to occur in terms of a function called *entropy*. Let us consider a reversible process such as the one shown in fig. 1. The following quantity

$$\int_A^B \frac{\delta Q}{T} \tag{1.3}$$

is the total heat transferred from the environment to the system during the reversible process. Notice that δQ is the heat transferred reversibly from the environment to the system. There are many possible reversible processes that drive the system from the initial equilirbium state A to the final equilibrium state B. It can be shown [Fer56] that this quantity does not depend on the (reversible) process and it can therefore be expressed as

$$\int_A^B \frac{\delta Q}{T} = S_B - S_A \tag{1.4}$$

that is, as a difference between the value of a function in the end of the process minus the value of the same function in the beginning of the process. That function is called *entropy* of the system, S. Notice that Eq. (1.4) is valid when the process is done reversibly. When the same process starting in equilibrium state A and ending in equilibrium state B is done irreversibly, and the integral in (1.4) is calculated along the process, the following inequality holds

Second Law of Thermodynamics

$$\int_A^B \frac{\delta Q}{T} \leqslant S_B - S_A, \tag{1.5}$$

where equal sign holds only for reversible processes. The above result is known as the *Second Law of Thermodynamics*

Let us consider an irreversible isothermal process in which a system is in contact with a single thermal environment that is at a constant temperature T. In this case, the integral in Eq. (1.5)

equals to the heat transfer over the temperature $\int_A^B \frac{\delta Q}{T}$. The entropy change of the system can be expressed as

$$\Delta S = \frac{Q}{T} + S_{prod}, \tag{1.6}$$

where S_{prod} is called the *entropy production* of the system. We can identify in Eq. (1.6) two terms that contribute to the change of entropy in the system. The first $\frac{Q}{T}$ is the entropy change due to the interaction of the system with the environment. The second, S_{prod} is part of the entropy change that is due to internal degrees of freedom of the system. The second law of thermodynamics (1.5) can be expressed in terms of the entropy production

$$S_{prod} \geqslant 0, \tag{1.7}$$

that is, the entropy production along a thermodynamic process increases. Another formulation of the second law can be done using the notion of the total entropy of the *universe*, the latter defined as the system plus the environment. In isothermal processes with a single heat reservoir, the change in entropy in the environment is equal to $-Q/T$ since the heat transferred *to* the reservoir has negative sign (the heat flows from the heat reservoir to the system). By adding the entropy changes of the system $Q/T + S_{prod}$ and of the reservoir $-Q/T$ we obtain the total entropy change in the universe ΔS_{tot} that equals to S_{prod} in this case, and therefore,

$$\Delta S_{tot} \geqslant 0. \tag{1.8}$$

The above equation is the expression of the second law of thermodynamics for the total entropy production in the universe.

In the case in which the system is in contact with a single thermal reservoir at temperature T, Eq. (**??**) is equivalent to

$$\Delta S - \frac{Q}{T} \geqslant 0. \tag{1.9}$$

If we consider the first law (1.1) and we replace the heat by $Q = \Delta U - W$ in the above equation we arrive at

$$W - (\Delta U - T\Delta S) \geqslant 0. \tag{1.10}$$

The following thermodynamic function

Helmholtz Free Energy

$$F = U - TS, \tag{1.11}$$

is called *Helmholtz Free Energy* or just *Free Energy*. The physical meaning of the free energy as a thermodynamic function can be done following Eq. (1.10),

$$W - \Delta F \geqslant 0. \tag{1.12}$$

The equal sign holds for reversible processes, $W_{rev} = \Delta F$, whereas for irreversible processes the work is $W_{rev} \geqslant \Delta F$. Therefore, for a process in which the system starts in an equilibrium state A and ends in another equilibrium state B, the free energy difference between the final and the initial state $\Delta F = F_B - F_A$ is the minimum amount of work that the external agent needs to perform to bring the system from the state A to the state B, $\Delta F = W_{rev}$. The *work dissipation* W_{diss} (or *dissipation*) is defined as the excess work needed to perform a physical process,

$$W_{diss} = W - \Delta F. \tag{1.13}$$

The dissipation is always positive and zero if the process is reversible,

$$W_{diss} \geqslant 0. \tag{1.14}$$

We have seen that the second law of thermodynamics imposes restrictions to some processes to occur. For example, if we put in contact two gases at different temperatures, one at T_1 and the other at $T_2 < T_1$ the mixture of the gases will reach an intermediate temperature T_{fin} whose value is between the initial temperatures of the gases, $T_1 < T_{fin} < T_2$. The second law forbids that. after the mixture, the temperature of the colder gas decreases and the temperature of the hotter gas increases.

1.2 MAXWELL'S DEMON AND SZILARD ENGINE

In 1871, Maxwell designed a thought experiment in which this situation was possible (see Fig. 3a) Imagine that the two gases are put in contact via a wall that can be open or closed at any time. Maxwell suggested that a "demon" *i.e.* an intelligent being able to see the molecules of each gas could in principle violate the second law of thermodynamics. The demon should control the wall in such a way that molecules that are colder than the average in the hot container (and vice versa) can pass through the wall. The result of the process is in principle forbidden by the second law: the colder gas gets colder and the hotter gas gets hotter. A simplified version of Maxwell's demon was introduced by Szilard [Szi29] in another thought experiment (see Fig. 3b). The *Szilard engine* is formed by a single particle confined in a container of volume V. The container is assumed to be immersed in a thermal reservoir of temperature T, therefore the particle thermalizes at temperature T after every collision with the walls of the container. After some time, a demon measures the position of the particle. If the particle is in the left half of the container, the demon introduces a movable piston in the right half of the container. On the other hand, if the particle is found on the

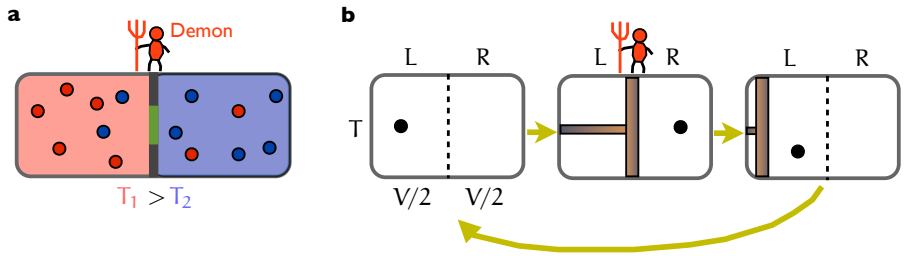


Figure 3: Maxwell's demon and the Szilard engine. **a.** Illustration of Maxwell's demon. Two gases are confined in separated containers and put in contact via a wall indicated in green in the figure. The gas on the left container is at a higher temperature T_1 than the gas on the right container, T_2. A demon can see the individual molecules of the gas and selectively open or close the wall at any time. If he opens the door so as he allows the slow molecules in the left container (in blue) to pass from left to right, and the fast molecules in the right container (in red) to pass from right to left, the temperature of the gas in the left container increases and the temperature of the gas in the right container decreases. **b.** Illustration of the Szilard engine. A single molecule is confined in a container that is immersed in a thermal bath at temperature T. The particle moves randomly along the two different halves of the container, L (left) and R (right). A demon measures where the particle is and introduces a movable piston in the right half if the particle is in the left half and vice versa. As a result, the particle collides against the piston until it reaches the wall of the container, which coincides with the initial condition of the cycle.

right half, the demon introduces the movable piston in the left half. After that, the particle collides against the piston until the piston reaches the boundary of the container. In this situation, the particle moves again along the whole container, that is, the system is back to the initial state.

The energetics of the Szilard engine plays an important role in the relationship between thermodynamics and information [Lef92]. We can assume that the energy needed to insert the piston is negligible since the insertion of the piston can be done reversibly [Paro1]. In principle, the demon is assumed to be small compared with the particle, therefore the energy spent by the demon to track the particle is also neglected. Under these two assumptions, the only energetic contribution comes from the expansion of the one particle gas against the piston from an initial volume $V/2$ to a final volume V. This energy can be considered as work performed by the system to the environment, and therefore it has to be negative. If we assume that the particle behaves like

a ideal gas of one particle, we can assume that the equation of state of the particle is $pV = kT$, the work done by the particle is

$$W = \int_{V/2}^{V} -p\,dV = -kT \int_{V/2}^{V} \frac{dV}{V} = -kT \ln 2. \tag{1.15}$$

Since the process is cyclic, $\Delta F = 0$, and since $W = -kT \ln 2 \simeq -0.69kT < 0$ along this process, the particle is systematically extracting work from the thermal reservoir, which is in contradiction with the second law (1.12).

This apparent contradiction with the second law of thermodynamics was solved by Landauer [Lan61]. Landauer investigated in this paper the connection between logical irreversibility and energy dissipation of computing processes. Since the information is always stored in physical devices, the nature of information is physical. Because of this, simple logical operations might have an energetic cost. In particular, the erasure of a bit is a logical irreversible operation, since a bit that is initially in one of the two possible states (0 or 1) is set to a reference state irrespective of its initial state. Landauer proved that the erasure of a bit requires *at least* $kT \ln 2$ of work dissipated to the thermal environment that is at temperature T,

$$W_{\text{eras}} \geqslant kT \ln 2. \tag{1.16}$$

Landauer's principle

This statement is known as *Landauer's principle*.

We can now reanalyze the energetics of the Szilard engine by taking into account Landauer's principle. In every cycle, there are two contributions to the total work: First, the work done in the system is the expansion work of the one-molecule gas. Second, contrary to what Szilard suggested, the work needed to measure where the particle is. Since the measurement has only two possible outcomes (L and R), the measurement in the Szilard engine is equivalent to a measurement of the state of a bit. In every cycle, the demon has to be restored to a reference state in order to be able to perform a measurement in the next cycle, which requires an amount of work that is at least $W_{\text{demon}} = W_{\text{eras}} \geqslant kT \ln 2$. The work per cycle of the Szilard engine is

$$W_{\text{cycle}} = W_{\text{gas}} + W_{\text{demon}} \geqslant -kT \ln 2 + kT \ln 2 = 0. \tag{1.17}$$

Notice that in the above proof we have assumed that a one particle gas behaves like a classical ideal gas. An alternative approach would be considering the particle as a quantum system. The "quantum version of the Szilard engine" was analyzed in [Zur03], where Szilard's result $W_{\text{cycle}} = kT \ln 2$ as well as Landauer's principle are recovered in the quantum limit.

1.3 REVIEW OF STATISTICAL MECHANICS

The connection between thermodynamic properties and the molecular dynamics of macroscopic physical systems was done with the introduction of statistical mechanics. Before that, the first attempts to explain equilibrium properties of macroscopic systems from the molecular motion, were done in kinetic theory. Boltzmann's theory aimed to describe the equilibrium properties of a gas in terms of the probability to find a molecule with velocity v at a given position x at time t, $\rho(x, v, t)$. One of the most remarkable results of Boltzmann's formalism is the analytic expression of the velocity distribution of the molecules of a gas in thermal equilibrium at temperature T. By assuming that there are only binary collisions between the molecules of the gas, Boltzmann found an expression for the velocity distribution that is independent on the interaction between molecules. In the simplest case, let us consider a gas in thermal equilibrium at temperature T formed by molecules of mass m. The gas is assumed to be in a one dimensional container. Under these assumptions, the probability to find a particle with velocity v at any position x and any time t is given by the *Maxwell-Boltzmann distribution*

$$\rho(v) = \sqrt{\frac{m}{2\pi kT}} e^{-mv^2/2kT}. \tag{1.18}$$

The probability distribution of the particle velocity is a Gaussian distribution with zero mean and dispersion

$$\langle v^2 \rangle = \frac{kT}{m}. \tag{1.19}$$

The above relationship connects with the *equipartition theorem*: In equilibrium, the average energy per degree of freedom of a physical system in contact with a thermal bath at temperature T is equal to $kT/2$. In this case, the only energetic contribution is kinetic. The average kinetic energy per molecule can be derived from (1.19),

Equipartition theorem

$$\langle E \rangle = \left\langle \frac{mv^2}{2} \right\rangle = \frac{kT}{2}. \tag{1.20}$$

Statistical mechanics goes further than kinetic theory and describes not only the temperature, but also other thermodynamic functions such as entropy or the free energy from the laws of molecular dynamics. Moreover, equilibrium properties of not only dilute gases but also of any macroscopic system can be studied within the framework of statistical mechanics [Hua87].

The first case studies of statistical mechanics are isolated systems. A physical system and the surrounding environment can be considered as a single bigger physical system (often called

"universe") whose total energy remains equal to a constant value E. The total number of particles N and the volume of the universe V is also considered as constant. The *macrostate* of the system is defined as the values of the thermodynamic parameters that characterize the state of the system, in this case (N, V, E). The *microstate* of the system is defined by the value of all the position and momenta of the molecules of the system, which in principle are three dimensional, $\{\vec{q}_i, \vec{p}_i\}$. The number of microstates that are compatible with a macrostate (N, V, E) is denoted by $\Omega(N, V, E)$. There are two postulates

- **Equiprobability**: All the microstates compatible with a given macrostate (N, V, E) are equally probable.
- **Equilibrium**: The equilibrium state is such that the number of compatible microstates $\Omega(N, V, E)$ is maximum.

Let us consider the situation in which two macroscopic systems that might be of different nature (gases, paramagnetic solids, etc.) are put in contact via a wall that allows the exchange of energy between them such as the example depicted in Fig. 4. The first

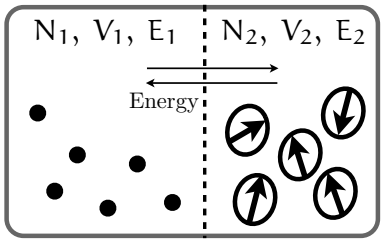


Figure 4: Energy exchange between two isolated systems. Two different subsystems, a gas (subsystem 1) and a paramagnetic solid (subsystem 2) in this case, are isolated from the environment. The energy exchange is allowed between the two subsystems but the number of particle and the volume of the two subsystems does not change in time. Despite allowing energy exchange between the two subsystems, the total energy is constant $E = E_1 + E_2$.

subsystem has N_1 particles, volume V_1 and its energy is labeled by E_1, whereas the second subsystem is characterized by N_2, V_2 and E_2. Since the total system is isolated, the total energy of the system does not change irrespective of the energy exchange, $E_1 + E_2 = E$. In this situation, the number of microstates of the total system compatible with a certain values of N, V and E is equal to

$$\Omega_T(N, V, E) = \Omega_1(N_1, V_1, E_1)\Omega_2(N_2, V_2, E_2), \tag{1.21}$$

since for every microstate of subsystem 1 there are Ω_2 microstates of 2 that are compatible with N, V, E. Because of the equilibrium

postulate of statistical mechanics, in equilibrium Ω_T is maximum, or equivalently, $d\Omega_T(N,V,E)/dE = 0$. This condition, together with the constraint $E_1 + E_2 = E$ yields the following condition

$$\frac{\partial \ln \Omega_1(N_1,V_1,E_1)}{\partial E_1} = \frac{\partial \ln \Omega_2(N_2,V_2,E_2)}{\partial E_2}. \tag{1.22}$$

We know from thermodynamics that the physical parameter that is equal in the two subsystems when they are put in contact via a diathermic wall is the temperature. Therefore, the quantity $\frac{\partial \ln \Omega(N,V,E)}{\partial E}$ has to be related with temperature. More precisely, the definition of temperature in statistical mechanics is

$$\frac{\partial \ln \Omega(E)}{\partial E} \equiv \frac{1}{kT}. \tag{1.23}$$

From Thermodynamics it is known [Cal85] that $\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V}$ means N, V constant. Equation (1.23) is consistent with the following definition of entropy,

Statistical definition of entropy

$$S(N,V,E) = k \ln \Omega(N,V,E), \tag{1.24}$$

which was introduced by Boltzmann. The above result is one of the first quantitative relationships done between a macroscopic quantity (a thermodynamic potential, S) with a microscopic property (the number of microstates. Ω) in statistical mechanics. Equation (1.24) shows that the entropy is larger when $\Omega(N,V,E)$ is larger, that is, when there are more states that are accessible by the system.

We notice that macroscopic parameters such as temperature are in general measured as a time average over a time window that is much larger than the atomic time scales. Statistical mechanics postulates that macroscopic parameters can also be obtained as averages over an *ensembles* of physical systems. From now, we will denote by (q,p) all the position and momenta of the system, *i.e.*, $(q,p) \equiv \{\vec{q}_i, \vec{p}_i\}_{i=1}^N$. We define by $\rho(q,p)$ the phase space density of the position of momenta of all the particles of the system, $i = 1, \cdots, N$ to be $\{\vec{q}_i, \vec{p}_i\}_{i=1}^N$. $\rho(q,p)dqdp$ is the probability of the system at time t to be in $[p, p+dp]$ and $[q, q+dq]$. The ergodic hypothesis assumes that for any thermodynamic function A,

$$\frac{1}{\Delta t}\int_t^{t+\Delta t} dt' A(t') = \int dqdp\ \rho(q,p)\ A(q,p). \tag{1.25}$$

The internal energy of the system can be obtained as the ensemble average of the Hamiltonian describing the system $H(q,p)$,

$$U = \langle H(q,p)\rangle = \int dqdp\ \rho(q,p) H(q,p). \tag{1.26}$$

Another particular case of thermodynamic function that can be expressed as an ensemble average is the entropy. Gibbs proved that entropy is the average of $-k \ln \rho(q,p)$, [Gre95]

Entropy (Gibbs formula)

$$S = \langle -k \ln \rho(q,p) \rangle = -k \int dq dp \; \rho(q,p) \ln \rho(q,p). \tag{1.27}$$

This result is consistent with Boltzmann's definition of entropy (1.24). Notice that the r.h.s. of (1.27) is proportional to a quantity that is commonly used in information theory *Shannon entropy* H, $H(\rho(x)) = \int dx \rho(x) \ln \rho(x)$, which is a measure of the uncertainty of the variable x distributed by $\rho(x)$ [Cov06].

Liouville's theorem

The time evolution of the phase space density $\rho = \rho(q,p;t)$ is governed by *Liouville's theorem*, [Pat88]

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{\rho, H\} = 0, \tag{1.28}$$

where $\{\cdot,\cdot\}$ denotes the *Poisson bracket* and

$$\{\rho, H\} = \sum_i \left(\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \tag{1.29}$$

$$= \sum_i \left(\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right). \tag{1.30}$$

Here i denotes the sum over all the degrees of freedom of the system. Liouville's theorem implies that, for an observer that moves in time, the density of representative points in phase space is conserved. An ensemble of representative points in phase space move like an incompressible fluid in physical space. An example of phase space densities that satisfy Liouville's theorem are those which depend on (q,p) via a Hamiltonian function $H(q,p)$, that is $\rho(q,p) = \rho(H(q,p))$. Depending on the boundary conditions (isothermal, isolated processes) the phase space density compatible with this constraint takes different forms as we now discuss.

For isolated systems described by a Hamiltonian $H(q,p)$, the phase space density is described by the *microcanonical ensemble*. Assuming that all the microstates that are compatible with a total energy E are equiprobable, the phase space density for all of them is equal to

$$\rho(\{q_i, p_i\}) = C\delta(H(q,p) - E), \tag{1.31}$$

where C is the normalization constant. The normalization of the phase space density in the microcanonical ensemble is equal to the inverse of the number of microstates compatible with a total energy equal to E,

$$C = \left[\int dq dp \; \delta(H(q,p) - E) \right]^{-1} = \frac{1}{\Omega(E)}. \tag{1.32}$$

In practical situations, the interactions between the system and the surroundings cannot be neglected, and the isolated systems are only an ideal case that are useful for calculation. We assume

isothermal conditions: The system exchanges energy with a thermal bath that does not change its temperature T even when exchanging energy with the system. In thermal equilibrium, the phase space density for the system is described by the *canonical* distribution,

Canonical distribution

$$\rho(q,p) = \frac{e^{-\beta H(q,p)}}{Z}. \tag{1.33}$$

where $\beta = 1/kT$, and the normalization constant

$$Z = \int dqdp e^{-\beta H(q,p)}, \tag{1.34}$$

is called the *partition function* of the system. The partition function contains all the thermodynamic information of the system. For example, the average energy of the system can be obtained from the partition function,

$$\langle H \rangle = -\frac{\partial \ln Z}{\partial \beta}. \tag{1.35}$$

A similar expression for the free energy in terms of the partition function can be found,

Free energy (canonical ensemble)

$$F = -kT \ln Z. \tag{1.36}$$

Statistical mechanics can be used to describe not only physical properties of macroscopic systems in equilibrium but also processes where the system is driven out of equilibrium [Zwa01, DG11, Bal75]. The driving is in general modeled by the change by the external agent of a control parameter λ according to a given protocol $\{\lambda(t)\}_{t=0}^{\tau}$. In a physical process, the variation of the average energy of a system can be expressed as the sum of two terms

$$\delta\langle H \rangle = \int dqdp\ \delta H\ \rho(q,p;t) + \int dqdp\ H\ \delta\rho(q,p;t), \tag{1.37}$$

where the chain rule was used inside the ensemble average. The first term is the variation of the average energy due to the variation of the hamiltonian, which can be related to the work done by the external agent. The second term only takes into account the variations of the phase space density, and it is related to the heat trasferred to the system. The work done by the external agent on a system defined by a Hamiltonian $H(q,p;\lambda)$ whose trajectory in phase space is $(q(t),p(t))$ and passes at time t through $(q(t),p(t)) = (q,p)$, $W(q,p;t)$, is given by

$$W(q,p;t) = \int_{\lambda(0)}^{\lambda(\tau)} \left.\frac{\partial H(q,p;\lambda)}{\partial \lambda}\right|_{q=q(t),p=p(t),\lambda=\lambda(t)} \cdot d\lambda(t). \tag{1.38}$$

If the system is in contact with a thermal bath at temperature T, and the process is isothermal and reversible, the probability

distribution of the system is canonical equilibrium distribution at any time $\rho(q,p;\lambda) = \frac{e^{-\beta H(q,p;\lambda)}}{Z_\lambda}$, where $Z_\lambda = \int dqdp e^{-\beta H(q,p;\lambda)}$, the average work $\langle W \rangle = \int dqdp\ \rho(q,p;t)\ W(q,p;t)$ yields the expected result, $\langle W \rangle = \Delta F$.

1.4 STOCHASTIC THERMODYNAMICS I: STOCHASTIC ENERGETICS

As we have shown in the previous section, classical thermodynamics applies only to systems with a large number of degrees of freedom and it establishes laws for macroscopic observables. When the size of the physical system under consideration is microscopic (or smaller) and the physical properties of the system can be sampled in a milisecond (or less) time scale, the dynamics of the system can be a priori described with the equations of motion of few degrees of freedom. The paradigmatic example is the motion of a microscopic object in a fluid at constant temperature, which is known as Brownian particle. Because of the small size of the particle, the thermal fluctuations of the environment and they make the motion of the particle to be random. Langevin equation [Lan08] provides a mesoscopic[1] description of the dynamics of the Brownian particle by describing the thermal fluctuations of the environment with a stochastic force that takes into account the degrees of freedom of the thermal bath. Recently, it has been possible to extend the notions of heat and work to systems obeying a Langevin equation, and the first law [Sek10] and the second law of thermodynamics [Sei05] have been studied in the mesoscopic scale.

1.4.1 *Brownian motion and Langevin equation*

The *Brownian* particle is the paradigmatic case study of stochastic thermodynamics. In 1828 [Bro28], Brown observed in the microscope the motion of grains of pollen in water. Brown used transparent grains so he could track their position, observing that their motion was apparently random. He checked this effect in different materials and claimed that this random motion has some connection to life but he did not made any connection to the random fluctuations of the environment. In 1905, Einstein [Ein05] proved that the erratic motion of such colloidal particles is due to random impacts of water molecules of the surrounding. He first developed a theory for an ensemble of Brownian particles moving in one dimension, showing that their average position

1 The *mesoscopic* description lies between micro and macroscopic description and consists on describing the motion of a microscopic system with few degrees of freedom

vanishes at any time $\langle x(t)\rangle = 0$, and that the average mean square displacement depends linearly on time

$$\langle x^2(t)\rangle = 2Dt, \tag{1.39}$$

being D the *diffusion coefficient*. The above formula is the basis of the diffusive motion, in which the distance travelled by the particles does not scale with time but with the square root of time. By using a simple model in which a balance between the gravitational force and the diffusion force is done, Einstein showed that the value of the diffusion coefficient in equilibrium is

Diffusion coefficient

$$D = \frac{kT}{\gamma}, \tag{1.40}$$

where k is Boltzmann's constant, T is the temperature of the surrounding fluid and γ is the friction coefficient of the particle in the fluid. In the case of an spherical particle of radius R immersed in a fluid of viscosity η, the value of the friction coefficient for laminar flow (low Reynolds number) is given by Stokes' law [Batoo] $\gamma = 6\pi\eta R$.

Eintein's relation (1.40) is a particular case of the *fluctuation-dissipation relationship* (FDR) that is introduced in linear response theory [Kub66]. Suppose that a system described by a single degree of freedom x is weakly perturbed from equilibrium by a force that depends on time, $f(t)$, which is applied from $t = 0$ where the system is in equilibrium. If the system is not perturbed far away from equilibrium, the average value of x at any time t depends linearly on the applied force via the *susceptibility* χ_x,

$$\langle x(t)\rangle = \int_0^t dt'\, \chi_x(t-t')f(t'). \tag{1.41}$$

The fluctuations of the variable x in time can be measured by the autocorrelation function,

$$C_x(t) = \langle x(t)x(0)\rangle. \tag{1.42}$$

Close to equilibrium, the following relationship between the Fourier transform of the autocorrelation and the imaginary part of the susceptibility holds

Fluctuation-Dissipation Relationship

$$\tilde{C}_x(\omega) = \frac{2kT}{\omega}\mathrm{Im}\tilde{\chi}(\omega), \tag{1.43}$$

which is known as the fluctuation-dissipation relationship [Kub66]. In linear response, the FDR can be used as a test to check whether or not a system is equilibrium with the thermal environment. Einstein's relation (1.40) can be derived from the FDR for a free Brownian particle described by the Langevin equation, which we now discuss.

A theory of the motion of an individual free Brownian particle was first done by Langevin [Lano8]. Langevin described the dynamics of the free Brownian particle of mass m starting with usual Newtonian equations of motion for the position $x(t)$ and momentum $p(t)$ of the particle. The total force that the Brownian particle feels is expressed as the sum of two terms. The first is the viscous force that the surrounding fluid exerts on the particle which is assumed to be in accordance with Stokes' law $F(t) = -\gamma v(t)$, being $v(t) = p(t)/m$ the instantaneous velocity of the particle. The second component of the force is a stochastic term $\xi(t)$ that plays the role of the fluctuating force exerted to the Brownian particle by the molecules of the environment. By considering these two forces, the equations of motion of the Brownian particle are

$$\frac{dp}{dt} = -\gamma \frac{p}{m} + \xi(t), \tag{1.44}$$

$$\frac{dx}{dt} = \frac{p}{m}. \tag{1.45}$$

Here $\xi(t)$ is a *Gaussian white noise* (see appendix A.1) of zero mean and variance, that is, a random force that is Gaussian distributed, its average vanishes and it is delta correlated with intensity given by the diffusion coefficient (1.40),

$$\langle \xi(t) \rangle = 0, \tag{1.46}$$

$$\langle \xi(t)\xi(t') \rangle = 2D\delta(t - t'). \tag{1.47}$$

Notice that Langevin equation is valid when all the degrees of freedom (position and momentum) that describe the surrounding fluid are rapidly varying degrees of freedom in comparison with the position and momentum of the colloidal particle. This assumption is true when the mass of the Brownian particle is several orders of magnitude larger than the mass of the molecules of the environment. In [Zwa73], Zwanzig presents a model of Brownian motion that can be analytically solved under the assumption that the correlations of the degrees of freedom of the molecules of the surrounding fluid decay fast. We notice that Langevin's description of the dynamics of a Brownian particle lies between the macroscopic approach and a microscopic description containing all the degrees of freedom of the Brownian particle and the surrounding fluid. This approach is commonly known as a *mesoscopic* description of the Brownian motion.

If the time resolution in which one measures the position of the Brownian particle Δt is such that $\Delta t \gg m/\gamma$, Eqs. (1.44,1.45) can be replaced by the following equation

$$\gamma \dot{x} = \xi(t), \tag{1.48}$$

where $\xi(t)$ is the same as in the underdamped case. The above equation is known as the *overdamped Langevin equation* for a free

Brownian particle. It can be obtained as the overdamped limit of Eq. (1.44), that it, when the mass of the particle is smaller than $m \ll \gamma \Delta t$ or the friction coefficient is larger than $\gamma \gg \Delta t/m$. A typical example of a system that obeys Langevin equation is a microscopic sphere of polystyrene immersed in water. For an spherical polystyrene particle of radius $R \sim 1\mu m$, the mass of the particle is of order $m \sim pg$ and the friction coefficient in water ($\eta = 0.89$mPas), assuming Stokes' law, is around $\gamma = 6\pi\eta R \sim 10^{-9} \frac{Ns}{m}$. This yields a characteristic time $\tau_m \sim 1\mu s$ which is much shorter than the usual experimental time resolution $\Delta t \sim 10^{-3}$s, being the overdamped limit $\Delta t \gg m/\gamma$ valid in this case.

In the case in which external forces (different from the forces applied by the environment) are applied on the Brownian particle, an additional term modelling the external forces has to been taken into account. If the external force is controlled via a control parameter $\lambda(t)$ and it is caused by a potential energy $V(x,\lambda)$, the underdamped Langevin equations for the particle are

$$\frac{dp}{dt} = -\frac{\partial V(x,\lambda)}{\partial x} - \gamma \frac{p}{m} + \xi(t). \tag{1.49}$$

$$\frac{dx}{dt} = \frac{p}{m}. \tag{1.50}$$

Notice that in general, the total external force applied on the particle $F(x,\lambda)$ has two contributions: First, a term that derives from a conservative potential $\frac{-\partial V(x,\lambda)}{\partial x}$ and second, a term that models the application of forces directly on the particle $f(x,\lambda)$. Therefore, in general, we should take into account both contributions and replace $\frac{-\partial V(x,\lambda)}{\partial x}$ by $F(x,\lambda) = \frac{-\partial V(x,\lambda)}{\partial x} + f(x,\lambda)$. In the overdamped limit, Eqs.(1.49,1.50) yield the overdamped Langevin equation in the presence of external conservative forces is

Langevin equation in the overdamped limit

$$\gamma \frac{dx}{dt} = -\frac{\partial V(x,\lambda)}{\partial x} + \xi(t). \tag{1.51}$$

This equation describes the random motion of colloidal microscopic particles but it can also be applied to magnetic or electric systems (see *e.g.* [Cof04]). More recently, Langevin equation has been applied to model the dynamics of molecular motors [Bus01, Jül97a].

1.4.2 *First law of thermodynamics in the mesoscopic scale*

In [Sek98], Sekimoto presented a pioneering work that introduced the stochastic notions of heat and work for systems obeying the Langevin equation, providing a new framework for defining the first law in stochastic thermodynamics. Let us consider first the

Langevin equation as a force balance equation in which the sum of all forces acting on the particle are zero,

$$0 = -\frac{\partial V(x,\lambda)}{\partial x} - \gamma\frac{dx}{dt} + \xi(t). \tag{1.52}$$

In the force balance equation, there are two forces that are exerted by the environment on the particle, the friction force $-\gamma\frac{dx}{dt}$ and the force of the environment $\xi(t)$. Therefore $(-\gamma dx/dt + \xi(t)) \circ dx(t)$ is the energy transfer from the environment to the particle, that is, the heat transferred to the particle. Notice that we have used the *Stratonovich product* $\circ$ (see Appendix A.2) for reasons that we explain below [Sek10]. We therefore define

Microscopic definition of heat

$$\delta Q \equiv \left(-\gamma\frac{dx}{dt} + \xi(t)\right) \circ dx(t), \tag{1.53}$$

as the heat transferred to an overdamped Brownian particle when it is displaced a differential amount $dx(t)$. In the underdamped limit, the above result can be expressed as

$$\delta Q = \left(-\gamma\frac{p}{m} + \xi(t)\right) \circ dx(t) = \left(\frac{dp}{dt} + \frac{\partial V(x,\lambda)}{\partial x}\right) \circ dx(t), \tag{1.54}$$

where we have used Eq. (1.49) in the second equality. We now focus on the energy balance of the particle. First, we notice that, since we use the Stratonovich product the usual rules of calculus apply, and we can express the two terms of $\delta Q = \left(\frac{dp}{dt} + \frac{\partial V(x,\lambda)}{\partial x}\right) \circ dx(t)$ in the underdamped limit as follows

$$\frac{dp}{dt} \circ dx(t) = \frac{dp}{dt} \circ \frac{p}{m} dt = d\left(\frac{p^2}{2m}\right), \tag{1.55}$$

$$\frac{\partial V(x,\lambda)}{\partial x} \circ dx(t) = dV(x,\lambda) + \frac{\partial V(x,\lambda)}{\partial \lambda}. \tag{1.56}$$

By summing Eqs. (1.55) and (1.55), we obtain

$$\delta Q = d\left(\frac{p^2}{2m} + V(x,\lambda)\right) + \frac{\partial V(x,\lambda)}{\partial \lambda} d\lambda. \tag{1.57}$$

The first term in (1.57) is the internal energy of the particle (kinetic and potential) $U(x,\lambda) = \frac{p^2}{2m} + V(x,\lambda)$. Therefore, by defining the second term in (1.57) as the work done by the external agent on the particle when it moves a small amount $dx(t)$,

Microscopic definition of Work

$$\delta W \equiv \frac{\partial V(x,\lambda)}{\partial \lambda} d\lambda, \tag{1.58}$$

we recover the first law of thermodynamics (1.1) in the microscopic scale

$$dU = \delta Q + \delta W, \tag{1.59}$$

with δW and δQ defined by Eqs. (1.58) and (1.53), respectively. We notice that in the overdamped limit, the heat (1.53) can be expressed as

$$\delta Q \equiv \frac{\partial V(x,\lambda)}{\partial x} \circ dx(t). \tag{1.60}$$

In this case, the first law is equivalent to a chain rule for the internal energy, since $dU = (\partial U/\partial\lambda) \circ d\lambda + (\partial U/\partial x) \circ dx(t)$ being the first term the work and the second, the heat.

The path dependence of the work and heat is now revealed in stochastic thermodynamics. The total work done by the particle and the total heat transferred to the particle depend on the specific stochastic trajectory, they are expressed as functionals of the microscopic trajectory. The total heat and work transferred from $t = 0$ to $t = \tau$ given that the agent controls the state of the system via $\lambda(t)$ and that the trajectory of the particle is $x_\tau \equiv \{x(t)\}_{t=0}^{\tau}$ is obtained by integrating Eqs. (1.53,1.58) with respect to time,

$$Q[x_\tau] = \int_{x(0)}^{x(\tau)} \frac{\partial V(x,\lambda)}{\partial x}\bigg|_{(x,\lambda)=(x(t),\lambda(t))} \circ dx(t), \tag{1.61}$$

$$W[x_\tau] = \int_{\lambda(0)}^{\lambda(\tau)} \frac{\partial V(x,\lambda)}{\partial \lambda}\bigg|_{(x,\lambda)=(x(t),\lambda(t))} \circ d\lambda(t). \tag{1.62}$$

Notice that the expression for the work is analogous to the general formula of the work given in statistical mechanics (1.38).

For a Brownian particle described by the overdamped Langevin equation, the free energy change in an isothermal process is equal to the work done by the external agent when the process is done quasistatically, *i.e.* when the system is at any time described by the canonical ensemble,

$$\Delta F = \int_{\lambda(0)}^{\lambda(\tau)} \left\langle \frac{\partial V(x,\lambda)}{\partial \lambda} \right\rangle_{eq} \circ d\lambda(t) \tag{1.63}$$

As it is proved in [Sek10], the stochastic definition of work is consistent with the usual statement of the second law of thermodynamics when considering the ensemble average (1.12) $\langle W \rangle - \Delta F \geqslant 0$. A further analysis on the second law of thermodynamics in the mesoscopic scale is done in the next subsection.

1.4.3 *Second law of thermodynamics in the mesoscopic scale*

The generalization of the second law of thermodynamics to the mesoscopic scale can be done by extending the notion of entropy to systems where thermal fluctuations are relevant. Seifert [Sei05] introduced a "trajectory entropy" for systems obeying an overdamped Langevin equation. His definition was previously proposed by Crooks [Cro99] for stochastic Markovian systems [Gri92]

obeying the microscopic reversible condition, and it was extended to colloidal particles in the overdamped limit and discrete systems governed by a master equation in [Sei05]. Let us consider a colloidal particle that moves in one dimension x from $t = 0$ to $t = \tau$. We assume that the total force exerted on the particle is $F(x, \lambda)$. In the overdamped regime, the following Langevin equation describes the motion of the particle,

$$\gamma \dot{x} = F(x, \lambda) + \xi, \tag{1.64}$$

where $\langle \xi(t)\xi(t') \rangle = 2\gamma k T \delta(t - t')$. We define the mobility by $\mu \equiv 1/\gamma$, and the diffusion coefficient is $D = kT/\gamma$. The dynamics of the particle can be described with a differential equation for the probability density of the particle to be in position x at time t, $\rho(x, t)$, which is called the *Fokker-Planck equation* (see Appendix A.3). In this case, the Fokker-Planck equation for $\rho(x, t)$ reads

$$\partial_t \rho(x, t) = -\partial_x j(x, t) = -\partial_x (\mu F(x, \lambda) - D\partial_x)\rho(x, t). \tag{1.65}$$

The Shannon entropy associated to the position of the particle at time t can be interpreted as a nonequilibrium entropy:

$$S(t) = -k \int dx(t)\ \rho(x(t), t)\ \ln \rho(x(t), t) \tag{1.66}$$

which suggest the introduction of the following trajectory dependent entropy

$$S[x(t)] = -k \ln \rho(x(t), t). \tag{1.67}$$

With this definition, the Shannon entropy at any time t can be seen as the ensemble average of $S[x(t)]$, $S(t) = \langle S[x(t)] \rangle$.

Using the trajectory dependent entropy, one can recover a second law of thermodynamics in the mesoscopic scale. Let us consider the entropy production of the system, *i.e.* $\dot{S}[x(t)]$. The time derivative is

$$\begin{aligned} \frac{\dot{S}[x(t)]}{k} &= -\left.\frac{\partial_t \rho(x, t)}{\rho(x, t)}\right|_{x(t)} - \left.\frac{\partial_x \rho(x, t)}{\rho(x, t)}\right|_{x(t)} \dot{x} \\ &= -\left.\frac{\partial_t \rho(x, t)}{\rho(x, t)}\right|_{x(t)} + \left.\frac{j(x, t)}{D\rho(x, t)}\right|_{x(t)} \dot{x} \\ &\quad - \left.\frac{F(x, \lambda)}{kT}\right|_{x(t)} \dot{x}, \end{aligned} \tag{1.68}$$

where we have used (1.65) in the second equality. The last term in (1.68) can be identified as an entropy change in the reservoir $\dot{S}_{res} = -\dot{Q}/T$, being Q calculated following (1.53). Since $-\gamma\dot{x} + \xi = -F(x, \lambda)$ the heat transferred to the particle is $\dot{Q}/T =$

$-\left.\frac{F(x,\lambda)}{kT}\right|_{x(t)}\dot{x}$ and the entropy change in the reservoir $\dot{S}_{res} = \left.\frac{F(x,\lambda)}{kT}\right|_{x(t)}\dot{x}$. The entropy production rate is (1.6)

$$\dot{S}_{prod} = \dot{S} + \dot{S}_{res} = k\left.\frac{\partial_t \rho(x,t)}{\rho(x,t)}\right|_{x(t)} + k\left.\frac{j(x,t)}{D\rho(x,t)}\right|_{x(t)}\dot{x}. \quad (1.69)$$

When averaging over $\rho(x,t)$, the first term in (1.69) vanishes, yielding [Sei05]

Second law of thermodynamics (mesoscopic scale)

$$\langle \dot{S}_{prod}(t)\rangle = k\int dx \frac{j(x,t)^2}{D\rho(x,t)} \geqslant 0. \quad (1.70)$$

This result is the expression of the second law of thermodynamics for the entropy production in the mesoscopic scale [cf. Eq. (1.7)]. Equation (1.70) is equivalent to the relationship between entropy and currents derived in linear irreversible thermodynamics [Kon98]. In a situation where a system is weakly perturbed away from equilibrium, the entropy production can be expressed as a linear combination of products between fluxes (currents) and forces, the latter being proportional to the currents itself. Therefore close to equilibrium, the entropy production is a positively defined quadratic form of the currents, as we will discuss in Chapter 2.1.1. The equality in (1.70) is obtained in equilibrium, where the current vanishes $j(x,t) = 0$.

We have seen that thermodynamic laws hold when considering averages of thermodynamic functions over many realizations of a process. Because of the importance of thermal fluctuations at microscopic scale, fluctuations can drive the system to exhibit deviations from the mean behavior. The introduction of the *fluctuation theorems* has provided a theoretical framework to measure the probability of the deviations from the mean in mesoscopic systems.

1.5 STOCHASTIC THERMODYNAMICS II: FLUCTUATION THEOREMS

In mesoscopic systems, the importance of fluctuations can lead to rare events such as an entropy decrease in the universe or negative dissipation in a single realization. As it was seen in the previous section, thermodynamic laws only apply to ensemble averages in the mesoscopic scale. It is interesting to know the probability of thermodynamic functions to take values that are different from the average. Fluctuation theorems are exact relationships that express universal properties of the probability distribution $\rho(\Omega)$ for thermodynamic functionals of the position $\Omega[x(t)]$ [Sei12] such as work, heat or entropy. Unlike the first results of nonequilibrium statistical physics that are valid only

in near-equilibrium conditions (small external forces), fluctuation theorems are valid for systems that are perturbed arbitrarily away from equilibrium. During the last decade, several fluctuation theorems have been derived and tested in simulations and experiments. We review some of the most relevant results have been obtained in the context of fluctuation theorems in the last decade.

1.5.1 *Jarzynski's equality*

One of the first results concerning the probability distributions of thermodynamic functions in nonequilibrium processes was introduced by Jarzynski in Ref. [Jar97]. Let us consider the following process: A system is initially in an equilibrium state A. An external agent changes some external parameters on the system according to a given protocol $\lambda(t)$. At the end of the process, the system is allowed to relax to a final equilibrium state B. In both isolated or isothermal conditions, the work done by the external agent in this process (given by (1.62)) is stochastic and the average over all the possible realizations satisfies the following inequality $\langle W\rangle \geqslant \Delta F$ (1.12). Moreover, the following equality holds

Jarzynski's equality

$$\langle e^{-\beta W}\rangle = e^{-\beta \Delta F}, \tag{1.71}$$

which is known as *Jarzynski's equality*. Notice that, since e^x is a convex[2] function, it satisfies Jensen's inequality, $\langle e^x\rangle \geqslant e^{\langle x\rangle}$, therefore Jarzynski's equality implies the second law $\langle W\rangle \geqslant \Delta F$. The main novelty of Jarzynski's result is that we can estimate an equilibrium property, the free energy difference between two equilibrium states, from a nonequilibrium average,

$$\Delta F = -kT \ln\langle e^{-\beta W}\rangle. \tag{1.72}$$

This result is valid for any protocol followed by the control parameter and it holds for isolated systems and for systems coupled weakly to a single thermal bath. It has been proved analytically [Jar97], in simulations [Par03] and experiments [Col05] as we will discuss in Sec. 1.6.

1.5.2 *Crooks's theorem*

Let us consider a system that is driven out of equilibrium in a finite time process defined by the evolution of a control parameter $\lambda(t)$ that changes from $t = 0$ to $t = \tau$. We call this process *forward process* and denote it by F. Let us now consider the time-reversed or *backward* (B) process in which the system is driven in the

2 A convex function $f(x)$ is such that for any $x_1 \neq x_2$, $\frac{1}{2}(f(x_1) + f(x_2)) \geqslant f(\frac{x_1+x_2}{2})$.

time-reversed manner $\tilde{\lambda}(t) = \lambda(\tau - t)$. We also assume that the system is in contact with a single thermal bath of temperature T . Crooks [Cro99] proved that the following fluctuation theorem holds for any nonequilibrium driving protocol, *Crooks's theorem*

$$\frac{\rho_F(W)}{\rho_B(-W)} = e^{\beta(W-\Delta F)}. \tag{1.73}$$

Here $\rho_F(W)$ is the work distribution of the forward process and $\rho_B(-W)$ is the distribution of $-W$ in the reverse process. This result is valid if the dynamics of the process is stochastic, Markovian [Gri92], the entropy production (1.69) is odd under time reversal and the dynamics is *microscopically reversible*. Microscopic reversibility condition is expressed as follows [Cro98]: the probability to observe a trajectory $\{x(t)\}_{t=0}^{\tau}$ in the forward process is related to the probability to observe the time reversed trajectory $\{x(\tau - t)\}_{t=0}^{\tau}$ in the time reversed process:

$$\frac{\mathcal{P}_F\left[\{x(t)\}_{t=0}^{\tau}\right]}{\mathcal{P}_B\left[\{x(\tau-t)\}_{t=0}^{\tau}\right]} = e^{-\beta Q[\{x(t)\}_{t=0}^{\tau}]}, \tag{1.74}$$

where $Q[\{x(t)\}_{t=0}^{\tau}]$ is the heat transferred to the system from the reservoir when the trajectory of the system is $\{x(t)\}_{t=0}^{\tau}$. In particular, systems described by a Langevin equation or Metropolis Monte Carlo satisfy all these conditions. However, the condition of the entropy production to be odd under time reversal is more restrictive, but it is hold for three classes of systems: (i) Systems that are initially in a thermal equilibrium state. (ii) Systems that are in a time symmetric nonequilibrium steady state (*e.g.* a gas in a piston that is driven sinusoidally or a fluid under a constant shear). In a more recent work, the validity of Crooks's fluctuation theorem has also been proved for an example of an isolated system following Hamiltonian dynamics provided that the initial condition of both forward and backward processes is canonical at temperature T [Cleo6].

We remark that Jarzynski's equality is direct consequence of Crooks's theorem. The average $\langle e^{-\beta W}\rangle$ reads,

$$\langle e^{-\beta W}\rangle = \int dW \rho_F(W) e^{-\beta W} = \int dW \rho_B(-W) e^{-\beta \Delta F} = e^{-\beta \Delta F}, \tag{1.75}$$

where the second equality is consequence of (1.73). Like Jarzynski's equality, Crooks theorem can be used to estimate the free energy difference between two equilibrium states if the system starts and end in two different equilibrium states. Let $W^\star$ be the value of the work at which the distributions $\rho_F(W)$ and $\rho_B(-W)$ coincide, $\rho_F(W^\star) = \rho_B(-W^\star)$. By means of (1.73),

$$\frac{\rho_F(W^\star)}{\rho_B(-W^\star)} = 1 \Rightarrow W^\star = \Delta F. \tag{1.76}$$

The above equation implies that we can also calculate the free energy difference between two equilibrium states from the intersection of the work distributions of any forward and time reversed nonequilibrium process.

1.5.3 *Steady state and transient fluctuation theorems*

Let us now consider a process that is driven to a nonequilibrium steady state (NESS) due to nonequilibrium constraints. One example is a system in contact with two thermal baths at different temperatures, such as a metallic bar with its two ends connected to baths at different temperatures. After some transient, the bar reaches a NESS where heat flows through the bar (it is out of equilibrium) from the hot to the cold reservoir and the temperature distribution does not change in time (it is stationary). Another example is a fluid between two parallel walls which move relative to each other at constant velocity and produce a constant shear on the fluid. In such processes, one cannot define a protocol $\lambda(t)$ since the nonequilibrium constraints do not change along the process. The NESS ensemble of the system is identical to the ensemble that is reached with the time reversed process but with a simple reflection. In these specific examples, the entropy production of the system (and also the total entropy production) is odd under time reversal and Crooks's theorem is valid. Since forward and time reversal processes are the same, we can drop off the index F, B and use ρ for distributions of the process in the NESS. Let us consider the observable defined by the time average of the entropy production rate $\dot{S}_{\text{prod}}$ in the interval $t \in [0, \tau]$ given by $S_\tau \equiv \frac{1}{\tau}\int_0^\tau \dot{S}_{\text{prod}}(t)dt$, where $t = 0$ denotes an arbitrary initial time of an interval of length τ in the steady state. In the NESS, the following fluctuation theorem holds

Steady state fluctuation theorem for entropy production

$$\ln\left(\frac{\rho_\tau(\dot{S})}{\rho_\tau(-\dot{S})}\right) = \frac{\dot{S}\tau}{k} \tag{1.77}$$

where $\rho_\tau(\dot{S})$ is the probability to observe a time average of the entropy production rate (in the stationary state) of value $S_\tau = \dot{S}$. Equation (1.77) is known as the *steady state fluctuation theorem* (SSFT) for entropy production. Entropy consumption in small systems for suitably short times was first observed in numerical simulations of a hard-disk fluid under a constant shear stress [Eva93], where the entropy production was found to be negative in several (but not the majority) of the realizations of the model (see Fig. 5). The SSFT is often called the *Gallavotti-Cohen theorem* since it was first formally proved by Gallavotti and Cohen for fluids under shear stress in which the chaotic

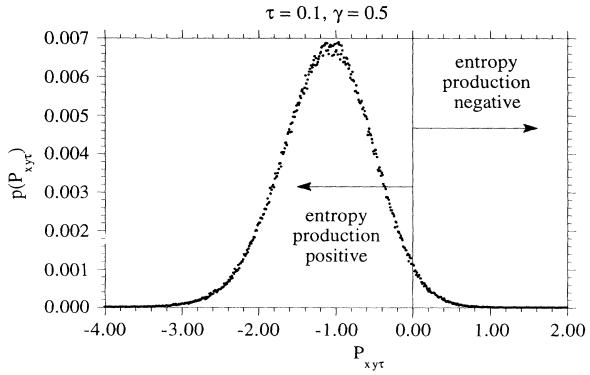


Figure 5: Entropy fluctuations in a hard-disk fluid under shear stress. The probability distribution of averaged stress tensor of a hard-disk fluid under shear stress is plotted as a function of the value of the stress tensor. Numerical simulations performed in [Eva93] for a shear rate $\gamma = 0.5$ and the time average of the stress tensor was done with $\tau = 0.1$. The vertical line separates data in which entropy was produced or consumed by the system. Picture taken from [Eva93].

molecular hypothesis[3] holds [Gal95]. Steady-state fluctuation theorem is only true in the long time limit [Eva96]. When τ is not very large compared to the transient time of the system, the left hand side of (1.77) is also a linear function of the entropy but the slope is not equal to $1/k$. In the long time limit, (1.77) can be rewritten using the entropy production rate, defined by $\dot{S} \equiv \lim_{\tau\to\infty} S_\tau = \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau \dot{S}_{\text{prod}}(t)dt$,

Steady state fluctuation theorem for entropy production (long time limit)

$$\frac{\dot{S}}{k} = \lim_{\tau\to\infty} \frac{1}{\tau} \ln\left(\frac{\rho_\tau(S)}{\rho_\tau(-S)}\right), \tag{1.78}$$

where $\rho_\tau(S)$ is the probability to observe an entropy production of value $S_{\text{prod}} = S$ in the steady state in the interval $t \in [0,\tau]$. Equation (1.78) is obtained as the limit of (1.77) when $\tau \to \infty$. The SSFT (1.78) is valid in both linear and non linear regime [Sea99] and also applies for systems obeying a Langevin equation [Kur99], for more general stochastic dynamics [Mae99] and for general Markov processes [Leb99]. In [VZ03b, VZ03a] the SSFT for both work and heat is found to be true for long τ for dragged Brownian particles that are described by the overdamped Langevin equation.

Notice that the SSFT predicts that in the NESS it is possible to observe negative entropy production in both finite time processes and also in the long time limit. However, the probability to observe a negative entropy production $-S$ is exponentially

3 *Molecular chaos hypothesis* holds that at any binary collision between two molecules the distribution of position and momenta of the two particles after the collision are independent each other.

smaller than the probability to observe a entropy production S, since $\rho(-S) = \rho(S)e^{-S/k}$. The second law of thermodynamics is recovered when averaging to the trajectory ensemble,

$$\frac{\langle \dot{S} \rangle}{k} = \lim_{\tau \to \infty} \frac{1}{\tau} \left\langle \ln \frac{\rho_\tau(S)}{\rho_\tau(-S)} \right\rangle, \tag{1.79}$$

where the right hand side is positive as we will see in Chapter 2, yielding the second law $\langle \dot{S} \rangle \geqslant 0$. Equation (1.79) also provides a connection between the average entropy production ($\dot{S}$) and a measure of time irreversibility of the process, as we will study in Chapter 2.

Notice that a *transient fluctuation theorem* (TFT) was introduced by Evans and Searles [Eva96, Sea99] by proving the same the same result as (1.77) but considering probability distributions of $S_\tau \equiv \frac{1}{\tau}\int_0^\tau \dot{S}_{prod}(t)dt$, $t = 0$ being the initial time of the process and obtaining $\rho(S_\tau)$ from an ensemble average over many trajectories. In [Eva96, Sea99] the TFT is proved for finite τ for isolated systems initially described by an equilibrium microcanonical ensemble that evolve towards a stationary state. Also in [VZ03b, VZ03a], the TFT for work and heat transferred to an overdamped Brownian particle is proved analytically to be true for any τ.

1.6 EXPERIMENTS

During the last years, the development of microscopic techniques and trapping techniques has allowed to study experimentally the thermodynamics of small systems and have experimental access to the energy fluctuations of a small system [Bus05]. We review some of the most relevant experiments done in small scale thermodynamics concerning the implementation of Szilard engine, the verification of Landauer's principle and the test of fluctuation theorems.

In many small-scale experiments, optical traps are the key experimental device since they are used to trap a single microscopic particle and to manipulate it, measuring forces of the order of piconewton and distances of the order of nanometer. *Optical tweezers* were first developed by Ashkin in the 80s. In 1986, Ashkin *et al* were able to trap dielectric particles in water with particle diameters ranging from 25nm to 10µm [Ash86]. Their experimental setup consisted on a single highly focused argon-laser beam (wavelength 541nm) as sketched in Fig. 6**a** that was used to trap dielectric beads immersed in a flow cell that contained water. A microscope was added to the setup to observe in real time the trapped particles as well as the diffraction pattern of the light. The trapping mechanism can be explained rigorously using Mie's theory [Mie08], which describes the interaction between

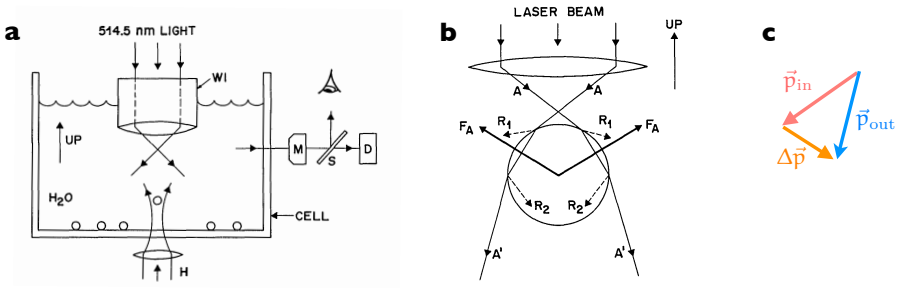


Figure 6: First single-beam optical trap. **a**. Sketch of the experimental setup. A 514.5nm argon laser light passes through a microscope that focuses the light in a cell where microscopic dielectric beads are immersed in water. A microscope (M) allows to observe the trapped particles. **b**. Ray optics of a spherical particle that is slightly after the optical focus of the microscope. The change in momentum of the light is indicated in the figure, as well as the direction of the reaction force of the bead (F_A). The net gradient force points to the focus. **c**. Momentum of the incident $\vec{p}_{in}$ and scattered $\vec{p}_{out}$ light and momentum change in the light $\Delta\vec{p}$ corresponding to the ray optics sketch of **b**. Pictures taken from [Ash86].

an electromagnetic plane wave and a dielectric sphere of a size comparable to the wavelength of the electromagnetic wave. Using ray optics we can describe the scattering and momentum transfer to the particle [Ash86] as we show in Fig. 6**b,c**. Because of third Newton's law the force exerted to the bead is opposite to the momentum change in the light. The net force in Fig. 6**b** points to the focus of the microscope, that is, in the direction of the gradient of intensity. This *gradient force* is compensated by the *scattering force* in the equilibrium position of the particle in the trap, which lies slightly after the focus of the microscope. In general the instensity profile of the laser is Gaussian. Because of this, when the particle moves perpendicular to the propagation axis, the light exerts a force on the bead pushing the bead towards the propagation axis as shown in Fig. 7. As a result, the particles are trapped in three dimensions by the optical tweezer as sketched in Fig. 7. The optical force exerted on the particle is Hookean (linear) for small deviations of the bead from the trap center, $\vec{F} = \vec{\kappa} \cdot (\vec{x} - \vec{x}_0)$, where $\vec{x}$ is the position of the bead, $\vec{x}_0$ is the equilibrium position of the bead in the trap and $\vec{\kappa} = (\kappa_x, \kappa_y, \kappa_z)$ is called *stiffness* of the trap.

The development of trapping techniques has allowed to build the first experimental microscopic Szilard engine formed by a single Brownian particle. In [Toy10], a microscopic particle is made to climb up a spiral-staircase-like potential created by an electric field using feedback control. A dimeric particle formed by two polystyrene spherical beads (both with diameter D = 287nm)

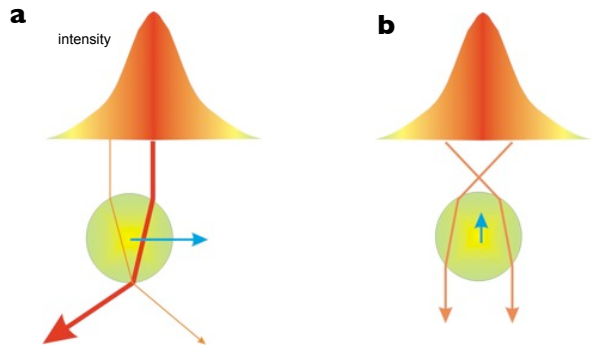


Figure 7: Optical forces in an optical tweezer. **a**. When the bead deviates from the propagation axis, the intensity profile is such that the bead pushes the light to the left and the light pushes the bead to the right, that is, through the propagation axis. **b**. The gradient force pushes the bead towards the focus of the microscope. Pictures taken from [Pet13].

is attached to the top glass surface of a fluid chamber. One of the particles of the dimer is held fixed and linked to the top surface of the chamber and the other performs rotational Brownian motion. Using elliptically rotating electric fields, they managed to induce a tilted periodic potential to the particle. The angular position of the rotating particle was measured with nanometer precision using video imaging. A feedback control loop of total time $t = \tau$ is implemented as follows: At $t = 0$, the angular position of the particle is measured. If the position lies within a specific interval, the phase of the potential is inverted at $t = \epsilon$ and the particle is let to relax to a new equilibrium position until $t = \tau$ in which the feedback cycle is repeated. When ϵ is comparable to the relaxation time of the particle in the periodic potential, the particle is made to climb upwards in the energy landscape.

More recently, Landauer's principle has been experimentally verified using optical tweezers [Bér12]. A single silica spheric bead (diameter $D = 2\mu m$) immersed in water was trapped using optical tweezers. By switching the laser position between two positions at a fast rate (faster than the relaxation time of the bead on the trap, τ_r), a bistable double-well optical potential was created. Berut *et al* were able to modulate the height of the barrier that separates the two minima of the double well potential by changing the intensity of the trapping laser. They designed a restore-to-zero cyclic process in which the bead is initially in one of the two equilibrium positions. Then, the barrier is lowered in 1s ($\gg \tau_r$) and the barrier is kept low for a time τ during which an external force that increases with time is applied. The cycle is finished lowering the barrier to its original height during 1s. This experimental protocol implements the erasure of a bit, since

the initial state of the particle ("left" or "right" well) is erased by changing the state of the bead to a reference state ("right" well in this case). In [Bér12] they measure the stochastic heat using (1.61) and show that in the limit of slow process ($\tau \simeq 40$s), Landauer bound (1.16) is experimentally checked, yielding $\langle Q \rangle \simeq kT \ln 2$ within experimental errors.

Thanks to the significant advance in trapping techniques, it has been possible to check deviations from the second law of thermodynamics in small systems [Bus05]. There have been several experimental verifications of the fluctuation theorems (Jarzynski's equality, Crooks's theorem, etc.) by manipulating systems where thermal fluctuations are relevant, such as DNA molecules or microscopic colloidal particles. We review some of the most relevant experiments done to check the fluctuation theorems introduced in Sec. 1.5.

- **Jarzynski's equality**: In Ref. [Lip02], Liphardt *et. al* test Jarzynski's equality with a single molecule experiment using RNA. They mechanically strech a RNA molecule between two different conformations. One of the ends of the RNA molecule is attached to a bead that is held fixed with an optical tweezers whereas the other end is attached to a second bead whose position can be controlled using a micropipette. The control parameter of the experiment is the distance between the two ends of the molecule. By measuring the force along the process, they are able to measure the stochastic work using (1.62). First, the free energy difference between the initial and final configurations is measured by stetching the RNA molecule reversibly, where $W \simeq \Delta F$. The same experiment is repeated irreversibly by moving the micropipette at a faster rate. By estimating the free energy difference using Jarzynski's equality (1.72), $\Delta F_{JE} = -kT \ln \langle e^{-\beta W} \rangle$, they find that the estimation from Jarzynski's equality is not significantly different from ΔF measured in the reversible process (the error is of order the thermal energy $kT/2$).

- **Crooks's fluctuation theorem**: With a similar setup as the one used to check Jarzinski's equality [Lip02], Collin *et. al* tested Crooks's fluctuation theorem (1.73) experimentally by unfolding and refolding a RNA hairpin [Col05]. A short interfering RNA hairpin which folds irreversibly but does not very far from equilibrium (dissipation of the unfolding is around 6kT) is stretched and unfolded at different pulling speeds. In Fig. 8a we show the RNA hairpin used in their experiment as well as the force extension curves (force applied as a function of the distance between the pulling ends). The pulling is done irreversibly, which makes the

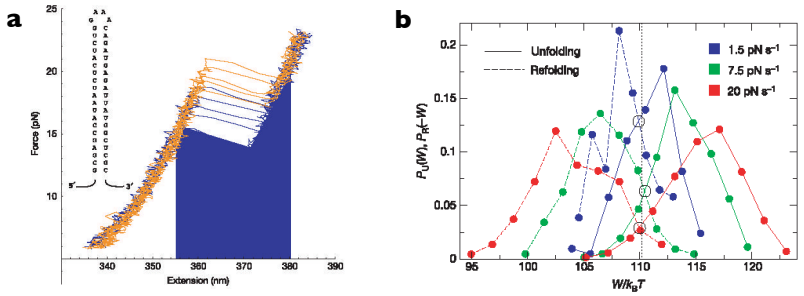


Figure 8: Experimental verification of Crooks's fluctuation theorem. **a**. RNA hairpin used in pulling experiments and five force extension curves of the unfolding (orange) and refolding (blue) curves at a load rate of 7.5pNs^{-1}. The shaded area is the work returned to the external agent in refolding processes. **b**. Unfolding (solid) $\rho_U(W)$ and refolding (dashed) $\rho_R(-W)$ work histograms at different pulling speeds. The distributions at different speeds cross around 110kT. Pictures taken from [Colo5].

force extension curves to exhibit hysteresis as shown in Fig. 8a. The area under the force extension curves of the unfolding (refolding) process is the work done along the unfolding (refolding) process. The distribution of the work of the unfolding $\rho_U(W)$ and refolding $\rho_R(-W)$ is shown in Fig. 8. As seen in Fig. 8b the work distributions $\rho_U(W)$ and $\rho_R(-W)$ cross at the same value of the work for different pulling speeds. The value of the work at which the two work distributions cross is equal to the free energy difference between the folded and unfolded states of the RNA hairpin measured with molecular dynamics simulations, reproducing Crooks's prediction (1.76).

- **Steady state (SSFT) and transient fluctuation theorems (TFT)**: The steady state fluctuation theorem (1.78) was first tested using a numerical simulation of a shearing fluid [Eva93] and later on, with a molecular dynamics simulation of a colloidal particle immersed in water translated by an optical trap [Wano2]. Since then, several experiments have been done in order to check the validity of both SSFT and TFT introduced in sec. 1.5.3. Ciliberto *et al* checked the SSFT (1.78) in the long time limit for the heat in a Rayleigh-Bernard cell made of water [Cil98] and also in turbulent flows [Cilo4]. Feitosa *et al* studied the energy fluctuations in a granular medium formed by 1mm-diameter spherical glass beads that are contained in a rectangular cage, finding that the SSFT holds in the long time limit [Feio4]. In [Garo5], the SSFT is verified for the power fluctuations in an electric circuit with a resistor in parallel with a capacitance. Schuler *et*

al [Sch05] studied the validity of the SSFT in a defect center in a diamond that is periodically excited by a laser, which is an example of a two-level system under a nonthermal noise. This experiment proves that fluctuation theorems also hold for nonlinear dynamics in which entropy fluctuations are not Gaussian. Jop *e*t al [Jop08] studied the fluctuations of the work and heat for a colloidal particle in an optical tweezer immersed in water. The optical tweezer is switched faster than the relaxation time of the bead to generate a double well potential, and the intensity is modulated sinusoidally, driving the particle to a steady state in which the SSFT for both work and heat is true in the long time limit. Wang *et al* [Wan05a] checked the TFT for a colloidal particle driven horizontally by an optical tweezer and the steady state fluctuation theorem when driving the tweezer circularly. In Ref. [GS10], the limitations of the SSFT are analyzed in two different scenarios: a colloidal particle trapped in a tweezer whose center moves stochastically and a atomic force microscopy cantilever close to a metallic surface.

1.7 OUTLINE AND GOALS

Understanding and exploiting the relationship between irreversibility and dissipation in nonequilibrium processes is one of the main goals of thermodynamics and statistical mechanics. Until the development of stochastic thermodynamics, the first quantitative study of this relationship was done in linear irreversible thermodynamics, where entropy production was found to be directly proportional to macroscopically observable flows or currents. This result, however, only applies to macroscopic systems that are weakly perturbed out of equilibrium and it does not explain the microscopic origin of the entropy production.

During the last years, stochastic thermodynamics has become an emergent field in physics, focusing on the understanding the dynamics and the energetics of systems of small scale. Fluctuation and work theorems have provided a theoretical framework that allows one to quantify the probability of rare events that only occur in the microscopic (or sub-microscopic) scale, such as the observation of negative entropy production in particular realizations of nonequilibrium processes. Recent results in this framework suggest that irreversibility and dissipation can be *quantitatively* connected for microscopic systems driven arbitrarily far from equilibrium.

This thesis is devoted to analyze the relationship between irreversibility and dissipation for microscopic systems that reach a nonequilibrium steady state (NESS). We are first interested in expressing this relationship in the NESS and secondly, we

aim to apply this result to practical situations. The main goal of this work is to find a technique that allows one to estimate the average dissipation of a microscopic system in a NESS from the statistics of a *single* stationary trajectory produced by the system during the process. We want our technique to be able to estimate the average dissipation of the system by using any data (one or several discrete or continuous degrees of freedom) sampled from the system and even ignoring any physical detail of the system. Our method uses the Kullback-Leibler divergence (KLD) to measure the distinguishability between a stationary trajectory and its time-reversal, providing a quantitative tool to measure the arrow of time in the NESS, *i.e.* the time irreversibility. We outline the content of the thesis according to the main goals of our research:

1. **Estimation of the KLD from single stationary trajectories.** In order to apply our technique to a variety of physical situations, we need to develop estimators of the KLD for both discrete and continuous data. We search for techniques that allow estimating the KLD using finite data from a single stationary trajectory, in order to apply our method to experimental data. This issue is addressed in Chapter 3.

2. **Application to discrete systems.** As a first case study, we will focus on discrete systems or coarse-grained descriptions of continuous systems. Using the existing techniques to estimate Shannon entropy of discrete data, we will extend this techniques to estimate KLD rate for discrete data obtained from a microscopic system in a NESS. We will check if we can distinguish between equilibrium and NESS by means of the value of the KLD and if we can infer or bound the dissipation of the physical mechanism that generated the discrete data. We will use a discrete ratchet model as a case study, where dissipation can be analytically calculated and compared to the KLD estimation using trajectories from simulations of the model. This goal is addressed in Chapter 4

3. **Application to continuous systems. Distinction between active and passive biological processes.** We extend the estimation technique to continuous data, in particular, to data produced by biological systems. More precisely, our goal is to use the KLD to distinguish between active and passive biological processes. Ear hair bundles are an excellent biological system to study this problem since previous works succeeded to detect active processes using data from spontaneous and forced oscillations of the bundles [Mar01b]. Our aim is to improve the method to discriminate between the active and passive case by applying our method only to

data from spontaneous oscillations. Both simulations and experimental data will be considered in this study. This issue is addressed in Chapter 5

4. **Study of the universality of the symmetry breaking.** We will show that the energetics of a generic symmetry breaking, as well the energetics of a symmetry restoration, can be reproduced with the relationship between irreversibility and dissipation for microscopic systems. We are interested in studying the relationship between entropy production and the probability to choose an instance in a symmetry breaking, *i.e.*, to study the thermodynamics of a choice of a generic probability p. This study can be done with the reverse process, that is, a symmetry restoration, which for a particular case yields Landauer's principle. This goal is addressed in Chapter 6

The energetics of symmetry breaking, as well as other interesting results of stochastic thermodynamics, can be experimentally tested using optical tweezers. Optical traps allow to measure physical properties of microscopic systems, where the energetics of the physical processes are of the order of the characteristic energy of thermal fluctuations, kT. In this thesis we are also interested in checking theoretical results in the framework of stochastic thermodynamics experimentally, using optical tweezers, which we now outline.

1. **Experimental check of the relationship between entropy production and probability to choose an instance in a process where a symmetry is broken.** Our aim is to design and implement an experiment that enables us to check the relationship between the probability to choose an option in a symmetry breaking with the entropy produced when a microscopic system chooses that option. A Brownian particle in a double-well potential with an external force that tunes the probability to stay in any of the two wells can be used as a case study. This situation might be implemented experimentally by trapping a Brownian microscopic particle using a dual optical tweezer. This goal is addressed in Chapter 6

2. **Temperature control in the microscopic scale.** In a recent work [Bli11], optical tweezers have been used to design microscopic heat engines, *i.e.* engines formed by a microscopic Brownian particle that operates between two thermal baths at different temperatures. We are interested in using optical tweezers to design an experimental technique to control the kinetic temperature of a Brownian particle and design

thermodynamic cycles in the spirit of [Bli11]. This issue is addressed in Chapter 7

Part II

IRREVERSIBILITY AND DISSIPATION

2

DISSIPATION AND KULLBACK-LEIBLER DIVERGENCE

In this chapter, we introduce the theoretical framework of the first part of our work, in which we study of the relationship between dissipation and irreversibility quantitatively in microscopic systems in the stationary state.

The relationship between entropy production (dissipation) and irreversibility forms the core of thermodynamics and statistical mechanics. The first studies in dissipation and irreversibility in nonequilibrium processes were done in the context of linear irreversible thermodynamics [Kon98]. In linear regime, the entropy production is linear with the force that drives the system out of equilibrium. However, this relation holds only in linear regime and does not provide a quantitative description of entropy production in terms of the microscopic properties of the system.

With the introduction of fluctuation theorems (see Chapter 1.5), it is possible to derive exact relationships that connect the entropy production of a microscopic system in the NESS with its microscopic properties and, moreover, provide a quantitative tool to measure the time irreversibility of the process. Our work is devoted to clarify this relation and to provide tools to estimate time irreversibility in the NESS from a single stationary trajectory [Rol10, Rol12, Lac12]. The theoretical framework of our approach is described in this chapter, whereas estimation techniques and applications to simulations and experimental data are described in further chapters. We will show how, given a stationary time series produced by a microscopic system, we can estimate the entropy production of the system that produced the series even ignoring any physical detail of the system. Our technique applies to both discrete and real-valued series as we will show in chapters 3,4 and 5.

This chapter is organized as follows: In Sec. 2.1 we review the notion of dissipation in irreversible processes from the approach of linear irreversible thermodynamics to the new insights provided by fluctuation theorems. In Sec. 2.2 we introduce the concept of relative entropy or Kullback-Leibler divergence and how it is related to the arrow of time. In Sec. 2.3 we show how dissipation and time irreversibility are quantitatively connected in the NESS. In Sec. 2.4 we show two examples of processes in which the relative entropy can be calculated analytically, Markov processes and hidden Markov chains.

2.1 AVERAGE DISSIPATIVE WORK IN IRREVERSIBLE PROCESSES

In nonequilibrium processes, the entropy production is a signature of the irreversibility of the process. In practical cases, such as a system in contact with a single heat bath at temperature T, the entropy production is revealed as an energy dissipation to the thermal bath $TS_{prod} = Q$. The first approach to the relation between irreversibility and dissipation was done in linear irreversible thermodynamics, where a process is said to be irreversible when macroscopic flows are observed in the system. The larger are the flows and therefore the irreversibility , the larger is the entropy production. However this approach is only valid in the linear regime. Using novel results derived in the context of fluctuation theorems we can extend the notions of irreversibility and dissipation not only to the linear regime but also to arbitrarily far from equilibrium processes that occur in the microscopic scale.

2.1.1 *Linear irreversible thermodynamics*

Linear irreversible thermodynamics studies systems that are driven not far from equilibrium in the context of linear response theory [Kon98]. Close to equilibrium, the entropy production of the system can be expressed as a linear combination of all the different thermodynamic forces or gradients F_i that are exerted on the system. The entropy production per unit volume, σ, can be expressed by

$$\sigma = \sum_i F_i J_i, \tag{2.1}$$

where i runs over all the different forces on the system and J_i is the flux associated to the force F_i. One example is the heat flux J_Q, which is produced by a force that is proportional to the gradient of temperature, $F_Q \propto \nabla \frac{1}{T}$. Another example is the electric current $J_e = I$ that is driven by an electric field E, $F_e \propto E$. In linear regime, the forces are proportional to the fluxes

$$F_j = \sum_k L_{jk} J_k \tag{2.2}$$

where Ljk are known as *phenomenological coefficients*. The above relation expresses that, for example, it is possible to induce a heat flow from an electric current, or vice versa. In this case, the entropy production per unit volume takes the following bilinear form in the flows

$$\sigma = \sum_{i,k} L_{ik} J_i J_k. \tag{2.3}$$

The second law of thermodynamcs, $\sigma > 0$, implies a constraint on the matrix $\mathbf{L} = L_{ij}$ such that it has to be positive definite. We notice that in this formulation, entropy production is positively defined quadratic form of the currents J and it is therefore related to the presence of macroscopically observable flows, since if $J > 0 \Rightarrow \sigma > 0$. This formulation however only applies to the linear regime and it does not connect the entropy production with the microscopic properties of the system. By using more recent results from fluctuation theorems, we can obtain a formula that expresses the entropy production of a microscopic system driven arbitrarily far from equilibrium and do a connection between the work dissipated and the microscopic properties of the system.

2.1.2 *Entropy production in microscopic systems*

As we showed in Chapter 1.4.3, the definition of the entropy associated to a trajectory of a microscopic system allows one to introduce the notion of entropy production in the case of a Brownian particle obeying an overdamped Langevin equation. By using the Fokker-Planck equation, the following expression relating ensemble average of entropy production and probability flux was derived by Seifert [Sei05],

$$\langle \dot{S}_{prod}(t) \rangle = k \int dx \frac{j(x,t)^2}{D\rho(x,t)} \geqslant 0. \tag{2.4}$$

Notice that this expresion was also obtained for Brownian chemical motors described by Langevin equation [Par02]. We notice that (2.1.2) expresses a relationship between entropy production and the probability current $j(x,t)$. According to , the entropy production a Brownian particle vanishes if $j(x,t)$ does, which means that irreversibility for these kind of systems is revealed in the flows of the system. Moreover, the entropy production depends on the current as j^2, which is in accordance with linear response theory, as shown in (2.3).

2.1.3 *KPB theorem*

Recently, a quantitative relationship between dissipation and irreversibility in nonequilibrium processes for microscopic systems has been derived. The introduction of fluctuation theorems has allowed to describe the entropy production of an isolated microscopic system with the microscopic details of the system [Kaw07, GM08b, Par09]. The main result was derived by Kawai *et al* in [Kaw07] and it is known in literature as the KPB (Kawai, Parrondo van den Broeck) theorem, which we now discuss.

Consider a physical system with Hamiltonian $H(q,p;\lambda)$, where (q,p) denotes a point in phase space, and λ is a parameter of the system controlled by an external agent. The system is initially isolated in equilibrium at temperature T, and the external agent modifies λ following a protocol $\{\lambda(t)\}_{t=0}^{\tau}$, with $\lambda(0) = \lambda_A$ and $\lambda(\tau) = \lambda_B$. We then let the system equilibrate by coupling it to a bath at temperature T'. The initial and final states of this process are equilibrium states for which entropy is well defined. We assume that the system is initially in $(q_0, p_0; 0)$ and ends in $(q_1, p_1; \tau)$ where q and p stand for the position and momentum of all the degrees of freedom of the system. The system is deterministic and the initial condition determines the whole trajectory. We denote by $W(q,p;t)$ the work associated to a process in which the system passes through (q,p) at time t. Since the system is isolated during the process

$$W(q,p;t) = H(q_1,p_1;\lambda_B) - H(q_0,p_0;\lambda_A). \tag{2.5}$$

We now try to relate the work done for a specific trajectory with the time irreversibility of the process by comparing the phase space densities of the process (*forward*) and the time reversed (*backward*) process, which is defined by the time-reversed protocol $\tilde{\lambda}_t = \lambda_{\tau-t}$ with $t \in [0,\tau]$. For a given trajectory in the forward process that starts in $(q_0, p_0; 0)$ and ends in $(q_1, p_1; \tau)$, the corresponding time reversed trajectory is obtained by reversing the position and changing the sign of the momenta. The time reversed trajectory starts in $(q_1, -p_1; \tau)$ and ends in $(q_0, -p_0; 0)$. Notice that the time t is taken in the forward process. We now compare the value of the phase space density of the forward process at an arbitrary time t measured in the forward protocol, $\rho(q,p;t)$, with the phase space density on the backward process at the *same* time t, $\tilde{\rho}(q,-p;t)$ as shown in Fig. 9. Because of Liou-

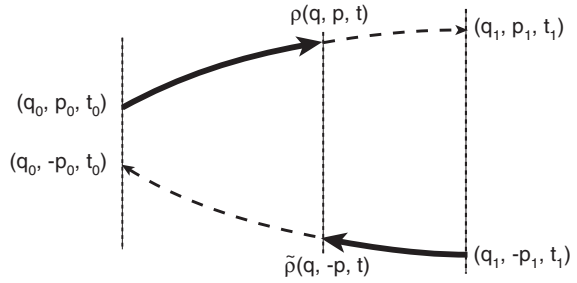


Figure 9: Forward and backward trajectories in phase space. Picture taken from [Kaw07].

ville's theorem 1.28, the phase space density is conserved in both forward and backward processes. Therefore, in both forward and

reverse processes ,the phase space density at any time coincides with its value in the beginning of the process,

$$\rho(q,p;t) = \rho(q_0,p_0;\lambda_A) = \frac{e^{-\beta H(q_0,p_0,\lambda_A)}}{Z_A}, \tag{2.6}$$

$$\widetilde{\rho}(q,-p;t) = \widetilde{\rho}(q_1,-p_1;\lambda_B) = \frac{e^{-\beta H(q_1,-p_1,\lambda_B)}}{Z_B}. \tag{2.7}$$

Where Z_A and Z_B represent the partition function in the initial and final (equilibrium) states, respectively. If the Hamiltonian of the system is even in momenta, $H(q_1,-p_1,\lambda_B) = H(q_1,p_1,\lambda_B)$, Eqs. (2.5,2.6) and (2.7) imply

$$\frac{\rho(q,p;t)}{\widetilde{\rho}(q,-p;t)} = e^{-\beta(W(q,p;t)-\Delta F)}, \tag{2.8}$$

where we have used $F = -kT\ln Z$ (1.36). The dissipative work reads,

$$W(q,p;t) - \Delta F = kT\ \ln\frac{\rho(q,p;t)}{\widetilde{\rho}(q,-p;t)}. \tag{2.9}$$

If we do the ensemble averages to all the possible trajectories in phase space of the forward process using $\rho(q,p;t)$, we arrive to the following result for the average dissipation *KPB theorem*

$$\langle W_{diss}\rangle = \langle W\rangle - \Delta F = kT\int dq\,dp\ \rho(q,p;t)\ln\frac{\rho(q,p;t)}{\widetilde{\rho}(q,-p;t)} \tag{2.10}$$

This result shows that the dissipation of a nonequilibrium process is revealed in the phase space. The right hand side is often called relative entropy or *Kullback-Leibler divergence* (KLD) [Kul51, Cov06] between the probability distributions $\rho(q,p;t)$ and $\widetilde{\rho}(q,-p;t)$,

$$\langle W_{diss}\rangle = kT\ D[\rho(q,p;t)||\widetilde{\rho}(q,-p;t)]. \tag{2.11}$$

As we will see in Chapter 2.2, D is a metric that compares two probability distributions. The KLD is positive, which ensures that the second law of thermodynamic holds in average $\langle W_{diss}\rangle = kT\ D[\rho(q,p;t)||\widetilde{\rho}(q,-p;t)] \geqslant 0$. The value of the KLD increases when the two probability distributions are more different each other, indicating that the more different are the forward and reverse process the more work is dissipated. The KLD between phase space densities in forward and backward processes is therefore a measure of the time irreversibility of a process, as we will see in Chapter 2.2.

In Fig. 10 we show an illustrative example of the applicability of KPB theorem. A system is initially in equilibrium at temperature T with an externally-controlled control parameter λ fixed at value λ_A. Then the system is disconnected from the thermal

bath and the control parameter is kept constant and equal to λ_A for a time longer than the relaxation time of the system, τ_r. Then the control parameter is changed from λ_A to λ_B linearly during a time τ in isolated conditions. In the end of the process, the control parameter is held fixed at λ_B for a time longer than τ_r and the system is let to relax by putting the system in contact with a thermal bath of temperature T. If the relaxation time of

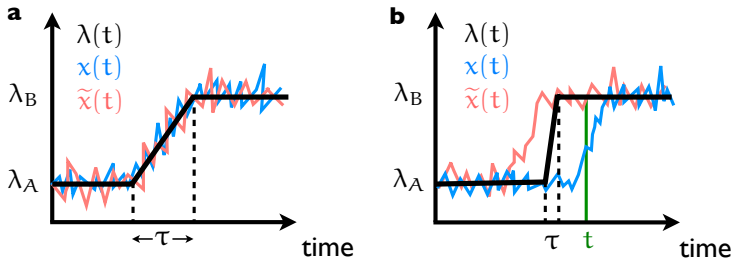


Figure 10: KPB theorem in an example. A system described by a single degree of freedom x is driven by an external agent following a protocol in which a control parameter is changed linearly from $\lambda(0) = \lambda_A$ to $\lambda(\tau) = \lambda_B$. **a**. Protocol (black line), and the value of x in the forward (blue) and backward process (red) when the total time of the process is much larger to the relaxation time of the system $\tau \gg \tau_r$. The dynamics is reversible and the phase space densities of forward and backward processes coincide at any time. **b**. Same graphs when the time of the process is much smaller than the relaxation time of the system, $\tau \ll \tau_r$. The dynamics is irreversible and the phase space densities of forward and backward do not coincide at any time. In this case, at time t indicated in green in the figure, $\rho(x,t) \neq \widetilde{\rho}(x,t)$ and we can measure the average dissipation of the forward process with $\langle W_{\text{diss}} \rangle = kTD[\rho(x,t)||\widetilde{\rho}(x,t)]$.

the system τ_r is very small compared to τ, $\tau_r \ll \tau$, the system relaxes to the equilibrium state at any time t and the process is done reversibly. In this case, the forward and reverse trajectories are indistinguishable and $\rho(x;t) \simeq \widetilde{\rho}(x;t)$ at any time t along the process. If the process is done much faster than the characteristic time scale of the system, $\tau_r \gg \tau$ then the system does not relax tho the equilibrium state during the process as shown in Fig. 10b. In this case, forward and backward trajectories are distinguishable, and in general $\rho(x;t) \neq \widetilde{\rho}(x;t)$. The KLD between these two distributions will be an estimation of the dissipation of the process.

The KPB theorem can be extended to a more general initial (equilibrium) conditions. In Ref. [Par09] it is proved that the change of the entropy in the system plus the bath, ΔS_{tot}, or

equivalently the entropy production on the system, averaged over many realizations of the process, satisfies

$$\langle \Delta S_{tot} \rangle = k\, D[\rho(q,p;t)||\tilde{\rho}(q,-p;t)]. \tag{2.12}$$

Equation (2.12) is valid for a variety of initial equilibrium conditions [Par09]: canonical, multi-canonical (several uncoupled systems at different temperatures), and grand-canonical distributions, as well as for different types of baths equilibrating the system at the end of the process. In all these cases, the formula is true when the evolution is isolated and the control parameter follows any arbitrary nonequilibrium protocol. In particular, for canonical initial conditions in the forward and in the backward processes, both at the same temperature T, Eq. (2.12) reads (see Ref. [Par09])

$$\begin{aligned} \langle \Delta S_{tot} \rangle &= \langle \Delta S_{system} \rangle + \langle \Delta S_{bath} \rangle \\ &= \frac{\langle \Delta U \rangle - \Delta F}{T} - \frac{\langle Q \rangle}{T} \\ &= \frac{\langle W \rangle - \Delta F}{T} = \frac{\langle W_{diss} \rangle}{T}, \end{aligned} \tag{2.13}$$

where Q is the heat exchanged with the thermal bath at the end of the process (realization dependent), and $W = \Delta U - Q$ is the work performed by the external agent. Therefore, in this specific case, entropy production equals the average dissipated work $\langle W_{diss} \rangle = \langle W \rangle - \Delta F$ divided by the temperature T and (2.12) becomes (2.11), $\langle W_{diss} \rangle = kT\, D[\rho(q,p;t)||\widetilde{\rho}(q,-p;t)]$.

2.2 KULLBACK-LEIBLER DIVERGENCE AND IRREVERSIBILITY

The Kullback-Leibler divergence (KLD), or relative entropy, measures the distinguishability of two probability distributions [Cov06]. Let us consider a random variable X and let p and q be two different probability distributions of the random variable X. We denote by $p(x)$ and $q(x)$ the probability of the variable X to take the value x when it is distributed according to p and q respectively. The KLD between the probability distributions p and q is defined by [Kul51]

Kullback-Leibler divergence (KLD)

$$D[p(x)||q(x)] = \int dx\, p(x) \ln \frac{p(x)}{q(x)}. \tag{2.14}$$

It is always positive, $D[p(x)||q(x)] \geqslant 0$ and vanishes if and only if $p(x) = q(x)$ for all x. Therefore, D follows two of the main properties of a mathematical *distance*. However, D is not symmetric with respect to a change of arguments, $D[p(x)||q(x)] \neq D[q(x)||p(x)]$ as it can be seen in the simple example of Fig.. An interpretation of $D[p||q]$ is the following: $D[p||q]$ is a measure of the amount of information needed to encode a sequence produced by q with a

code that is optimal for sequences generated by p. Interpretation of the example is as follows: it is easier to obtain with dice a sequence drawn with a lottery than the opposite.

The interpretation of the KLD as a measure of distinguishability is a consequence of the *Chernoff-Stein lemma* [Cov06]: the probability of incorrectly guessing (via hypothesis testing) that a sequence of n data is distributed according to p when the true distribution is q is asymptotically equal to $e^{-nD[p(x)||q(x)]}$. Therefore, when p and q are similar —in the sense that they overlap significantly— the likelihood of incorrectly guessing the distribution, p or q, is large [Cov06]. Chernoff-Stein lemma implies that the KLD in (2.11) can be considered as a measure of the arrow of time, since it measures the difficulty to distinguish whether the state (q,p) of the system at time t was generated in the forward on backward experiment [Par09].

We now remark that the relative entropy in KPB theorem (2.11),

$$D[\rho(q,p;t)||\widetilde{\rho}(q,-p;t)] = \int dq\,dp\ \rho(q,p;t)\ln\frac{\rho(q,p;t)}{\widetilde{\rho}(q,-p;t)}, \quad (2.15)$$

is not a proper KLD, since the argument of *the two* probability distributions is not the same as in the definition of the KLD (2.14) The KLD $D[\rho(q,p;t)||\widetilde{\rho}(q,-p;t)]$ has to be understood as a relative entropy between two probability distributions of the *same* random variables (q,p) at time t: $\rho(q,p;t)$, and a second distribution $\sigma(q,p;t)$.

$$D[\rho(q,p;t)||\sigma(q,p;t)] = \int dq\,dp\ \rho(q,p;t)\ln\frac{\rho(q,p;t)}{\sigma(q,p;t)}, \quad (2.16)$$

such that the value of $\sigma(q,p;t)$ for (q,p) satisfies $\sigma(q,p;t) = \widetilde{\rho}(q,-p;t)$.

Let us recall a property of the KLD that we use throughout our work [Cov06]. If we have two random variables X,Y and two joint probability distributions $p(x,y)$ and $q(x,y)$, then the *chain rule* holds (see Appendix B.1),

Chain rule of the Kullback-Leibler Divergence

$$D[p(x,y)||q(x,y)] \geqslant D[p(x)||q(x)]. \quad (2.17)$$

This means that it is harder to distinguish between p and q when we consider only the marginal distributions, $p(x)$ and $q(x)$, instead of the full joint distributions, $p(x,y)$ and $q(x,y)$. If X,Y describe the state of a system, Eq. (2.17) indicates that the KLD decreases when only a partial description of the system, given by the variable X, is available. The bound in (2.17) is an equality when the variable Y carries redundant information with respect to the variable X, for example, when Y is obtained as a function of X, $Y = f(X)$ for any function f.

The chain rule has an important consequences on the relationship between irreversibility and dissipation established by the

KPB theorem. If we do not have access to all the details of the system and the phase space can only be sampled partially, the KLD yields a lower bound to the dissipation. considering *partial information* that is when not all the degrees of freedom can be sampled. When not all the degrees of freedom of the system can be sampled, we say that *partial information* of the physical is available. Let x be *any* collection of m position and n momenta of the system $x = (q_1, \cdots, q_m; p_1, \cdots, p_n)$, where m and n can be different and $3n + 3m$ is smaller than the total number of degrees of freedom of the system. Since x describes in general only a part of the physical system we say that x contains *partial information* of the system. Because of the chain rule, when the state of the system is described with partial information x, the KPB theorem (2.11) turns into an inequality

Dissipation and KLD with partial information

$$\langle W_{\text{diss}} \rangle \geqslant kT\, D[\rho(x;t)||\widetilde{\rho}(\tilde{x};t)], \tag{2.18}$$

where $\tilde{x} = (q_1, \cdots, q_m; -p_1, \cdots, -p_n)$. We notice that even ignoring the full information of phase space (2.18) $D[\rho(x;t)||\widetilde{\rho}(\tilde{x};t)]$ still gives at least a lower bound to the average dissipation that is in accordance with the second law of thermodynamics, $\langle W_{\text{diss}} \rangle \geqslant kT\, D[\rho(x;t)||\widetilde{\rho}(\tilde{x};t)] \geqslant 0$. Consequently, when the system is described using only a reduced set of variables of the phase space, the KLD in (2.18) provides a lower bound to the dissipation. There are two groups of variables that provide *redundant* information to measure the irreversibility with the KLD. First, variables which are time reversible (their distribution is the same in the forward and reverse processes). Second, system variables that are obtained as a function of other variables. All these *redundant* variables do not provide any extra statistical information about the direction of the arrow of time of the system, and they can be ignored to measure the dissipation using the KLD [GM08b, Rol10, Rol12].

Now suppose that at every t, we cannot measure the micro state $(q, p; t)$ of the system but we can only detect that the system is in a specific subset of the phase space. We therefore describe the system using a *coarse-grained* random variable X that indicates in which subset the system is at every time t. For example, the position of a Brownian particle in one dimension x can be coarse-grained by introducing a new variable α that indicates if the position is positive or negative, for example $\alpha = 0$ if $x \leqslant 0$ and $\alpha = 1$ if $x > 0$. The KPB theorem is found to be an inequality when coarse-graining [Kaw07]. If we phase space is partitioned in K non overlapping subsets $\{\mathcal{X}_j\}_{j=1}^{K}$ the coarse-grained forward and backward phase space densities are

$$\rho_j(t) = \int_{\mathcal{X}_j} dq dp\ \rho(q,p;t); \quad \widetilde{\rho}_j(t) = \int_{\widetilde{\mathcal{X}}_j} dq dp\ \widetilde{\rho}(q,-p;t), \tag{2.19}$$

where $\widetilde{\rho}_j$ is identical to ρ_j except a change of sign in all the momenta. At an arbitrary time t during the nonequilibrium process the following coarse-grained KLD

Dissipation and KLD with coarse-grained information

$$D[\rho(t)||\widetilde{\rho}(t)] = \sum_{j=1}^{K} \rho_j(t) \ln \frac{\rho_j(t)}{\widetilde{\rho}_j(t)}, \tag{2.20}$$

is a lower bound to the average dissipation [Kaw07]

$$\langle W_{diss} \rangle \geqslant kT\, D[\rho(t)||\widetilde{\rho}(t)]. \tag{2.21}$$

This result indicates that the time irreversibility measured using a coarse-grained description of the micro state of the system provides a lower bound to the average dissipation of the process.

The chain rule allows one to rewrite the KPB theorem using the KLD between the forward and reverse distributions of *trajectories* in phase space. For isolated systems, the evolution is deterministic, except for the last stage where the system is connected to the bath, and the point $z = (q, p)$ at time t determines the whole trajectory of the system $\{z(t)\}_{t=0}^{\tau}$. Then $z(t)$ and $\{z(t)\}_{t=0}^{\tau}$ carry the same information and the KLD of their respective probability densities are equal by virtue of the chain rule. Equation (2.11) can be rewritten in terms of *path probabilities*. Let $\mathcal{P}(\{z(t)\}_{t=0}^{\tau})$ be the probability to observe a trajectory $\{z(t)\}_{t=0}^{\tau} = \{q(t), p(t); t\}_{t=0}^{\tau}$ in the forward process. The corresponding time-reversed path in the backward process starts in $(q(\tau), -p(\tau); \tau)$ and ends in $(q(0), -p(0); 0)$ as shown in Fig. 9. In a more compact notation, the time-reversed trajectory is defined as $\{\tilde{z}(\tau - t)\}_{t=0}^{\tau} = \{q(\tau - t), -p(\tau - t); \tau - t\}_{t=0}^{\tau}$. The probability to observe such a trajectory in the backward process is denoted by $\widetilde{\mathcal{P}}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})$. The KLD between forward and backward trajectory distributions equals to the average energy dissipated in the process [GM08b, Rol10, Rol12]

KPB theorem for trajectories

$$\langle W_{diss} \rangle = kT\, D[\mathcal{P}(\{z(t)\}_{t=0}^{\tau})||\widetilde{\mathcal{P}}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})]. \tag{2.22}$$

The above KLD has to be understood as a KLD between two distributions of a stochastic process [cf. (B.10) in Appendix B.1]. As noticed before in (2.16), the right hand side in (2.22) is a KLD between $\mathcal{P}$ and a different trajectory distribution $\mathcal{Q}$,

$$\langle W_{diss} \rangle = \int \mathcal{D}(\{z(t)\}_{t=0}^{\tau}) \mathcal{P}(\{z(t)\}_{t=0}^{\tau}) \ln \frac{\mathcal{P}(\{z(t)\}_{t=0}^{\tau})}{\mathcal{Q}(\{z(t)\}_{t=0}^{\tau})}, \tag{2.23}$$

where $\mathcal{Q}$ is such that $\mathcal{Q}(\{z(t)\}_{t=0}^{\tau}) = \mathcal{P}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})$.

We now recall that using Crooks's fluctuation theorem, we can arrive to an expression of the average dissipative work in terms of the forward and backward work distributions. Integrating Crooks's theorem (1.73), $W - \Delta F = \ln \frac{\rho(W)}{\widetilde{\rho}(-W)}$, where $\rho(W)$

$[\widetilde{\rho}(W)]$ is the probability density of the work done on the system along the forward (backward) process [Cro99, GMo8b], one immediately gets

$$\langle W_{\text{diss}} \rangle = kT\, D[\rho(W)||\widetilde{\rho}(-W)]. \tag{2.24}$$

Notice that the work W is a functional of the trajectory $\{z(t)\}_{t=0}^{\tau}$ [see (1.62)] containing much less information than the trajectory itself. As indicated by the chain rule (2.17), the KLD of work distributions should in principle be smaller than the KLD of trajectory distributions. On the contrary, the KLD is the same, indicating that all the irreversibility of the process is captured by the dissipative work [GMo8b]. Equation (2.24) indicates that the irreversibility in the work reveals the entropy production.

If the micro state of the system is not known are every time t but only a trajectory containing the evolution of a subset of variables of the phase space, $\{x(t)\}_{t=0}^{\tau}$, the KLD between forward and backward trajectories yields bounds the dissipation from below,

$$\langle W_{\text{diss}} \rangle \geqslant kT\, D[\mathcal{P}(\{x(t)\}_{t=0}^{\tau})||\widetilde{\mathcal{P}}(\{\tilde{x}(\tau - t)\}_{t=0}^{\tau})]. \tag{2.25}$$

The bound saturates when the variables that are sampled capture the same information about the irreversibility as the work. The information of the irreversibility is contained in the variables that interact with the work-performing device. These variables are the *footprints* of irreversibility [GMo8b].

A different situation in which partial information of the system is available occurs when the state of the system cannot be sampled at any time t along the process but only every finite time interval Δt. In this case, even if the micro state of the system $(q(t), p(t); t)$ is known every Δt, the KLD gives a lower bound to the dissipation. In [GMo8a], Gomez-Marin *et al* study the dynamics of an overdamped Brownian particle that moves in one dimension dragged by a harmonic potential that moves at constant velocity. They are able to measure the KLD between trajectories of the position of the particle $\{x(t + i\Delta t)\}_{i=0}^{\tau/\Delta t}$ in the forward process and the backward process, the latter consisting in moving the trap at the opposite velocity as in the forward process. When reducing the sampling time Δt, the KLD approaches asymptotically to the entropy production in the system.

In summary, when considering the KLD between forward and backward distributions of trajectories, the irreversibility is partially captured by the KLD when one or more of the following phenomena occur:

- **Coarse graining**: Only a coarse-grained description of the system is available, that is, we only know in which of an ensemble of phase space subsets the system is.

- **Partial information**: Not all the variables of phase space can be sampled but only a subset of variables in phase space.
- **Finite time sampling**: The trajectory cannot be sampled at any time but only at a finite sampling frequency.

When the information about the system is not full in the sense that one of the three above mentioned shortcomings occur, the KLD between forward and backward trajectories is a lower bound to the average entropy production in the system. In Fig. 11 we

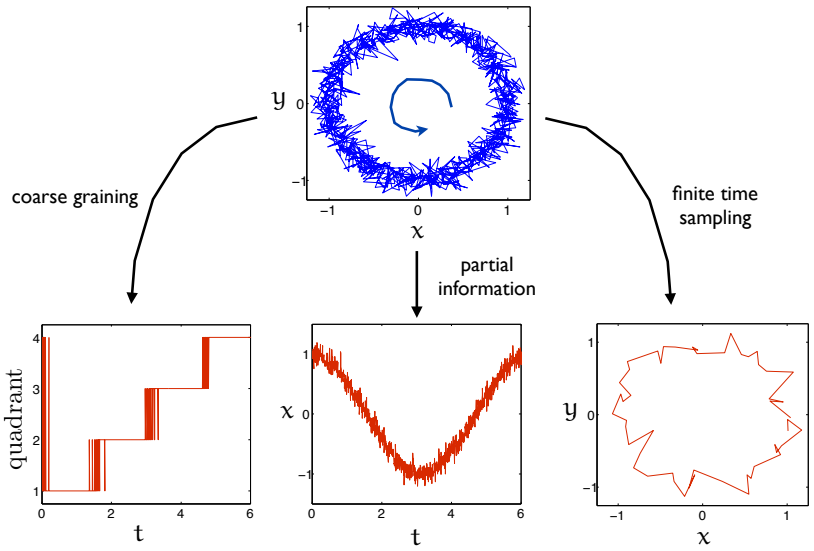


Figure 11: Illustrative example of possible information shortcomings. A stochastic oscillator is defined by the set of equations $x(t) = \cos(t) + 0.10 \cdot w_x(t)$; $y(t) = \sin(t) + 0.10 \cdot w_y(t)$, where w_x and w_y are random independent numbers distributed as $\mathcal{N}(0,1)$. The time runs from $t = 0$ to $t = \tau = 2\pi$ in steps of $\Delta t = \tau/1000$. We show the trajectory of the system in the $x - y$ plane and also when using partial descriptions: measuring the quadrant in which the system is at any time (bottom left panel), measuring only the variable x at any time (bottom center panel) and sampling the trajectory in $x - y$ plane every $20\Delta t$.

show an illustrative example where the three types of lack of information can occur. An stochastic oscillator is described by two degrees of freedom x and y. The trajectory in the $x - y$ plane, $\{x(t), y(t)\}$, is irreversible because the oscillator moves in average with a positive angular frequency. If we consider a variable q that measures in which quadrant the system is at any time we get a coarse grained description from the time series $\{q(t)\}$. When sampling the system with partial information given by the variable x, the trajectory $\{x(t)\}$ is time reversible as shown in Fig. 11 which corresponds with a vanishing KLD (and therefore

is a lower bound to the KLD calculated with full information). A finite time sampling of the original trajectory is also shown in Fig. 11. In any of these cases, the KLD decreases when part of the full information of the system is missing.

Irreversibility and dissipation in isothermal processes

Until now, we have only considered the relation between dissipation and irreversibility for isolated systems. However, the KPB theorem can be extended also to systems that are in contact with a thermal bath at temperature T. When a system is in contact with a thermal bath at temperature T, the system plus the bath can be viewed as an isolated "super system". The KPB theorem implies that the average dissipation of the system plus the bath equals to $\langle W_{\text{diss}} \rangle = kT\ D(\rho_F||\rho_B)$ where $D(\rho_F||\rho_B)$ is calculated in the full phase space (system plus bath). When considering only the system variables, we are neglecting the variables of the bath, and by virtue of the chain rule, the KLD yields a lower bound to the dissipation when considering only the phase space density of the system, $\langle W_{\text{diss}} \rangle \geqslant kT\ D(\rho_F||\rho_B)$. In [Hor09], this result is illustrated with an overdamped Brownian particle in a dragged harmonic trap immersed in a thermal bath at temperature T is studied. It is shown that in isothermal conditions, a lower bound to the dissipation is obtained using the KLD $\langle W_{\text{diss}} \rangle \geqslant kT\ D(\rho_F||\rho_B)$. The bound is tighter the smaller is the friction coefficient γ, and therefore the weaker is the coupling between the Brownian particle and the reservoir. When $\gamma \to 0$, the system is uncoupled to the bath and the equality is found.

2.3 DISSIPATION AND IRREVERSIBILITY IN THE NONEQUILIBRIUM STATIONARY STATE

The main goal of our work is to explore the existing quantitative relation between dissipation and irreversibility for microscopic systems for the case of nonequilibrium processes that reach a nonequilibrium stationary state (NESS). As we revised in Chapter 1.5.3, a fluid under a constant shear, a gas in a piston that is moved sinusoidally or a molecular motor driven by the ATP hydrolysis are only a few examples of nonequilibrium processes that reach a NESS.

We now proceed to apply the above results to stationary trajectories. Consider a long process in which the system reaches a nonequilibrium stationary state (NESS) after a possible initial transient. In the NESS the external parameter is held fixed, $\lambda(t) = \lambda$ or it is time-symmetric; the system is kept out of equilibrium due to the existence of baths at different temperatures (a possibility that is included in the hypothesis used in [Par09] to prove (2.12)) or different chemical potentials, external constant forces, etc. In the steady state, the protocol and its time reversal are identical $\lambda(t) = \lambda(\tau - t) = \tilde{\lambda}(t)$. Therefore the probability

Probability distributions in the NESS

distributions of the process and its time reversal are identical in the steady state,

$$\widetilde{\mathcal{P}}_{ss} = \mathcal{P}_{ss}. \tag{2.26}$$

For simplicity, we will denote by $\mathcal{P}$ the trajectory probability density in the stationary state. In the long time limit, $\tau \to \infty$, we can neglect the contribution of the transient to the entropy production and rewrite (2.12) for the entropy production per unit of time $\dot{S}$ in the NESS [Mae03] as

$$\langle \dot{S} \rangle = k \lim_{\tau\to\infty} \frac{1}{\tau} D\left[\mathcal{P}\left(\{z(t)\}_{t=0}^{\tau}\right) || \mathcal{P}\left(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau}\right)\right]. \tag{2.27}$$

As we showed in Chapter 1.5.3, a similar expression can be obtained from the steady state fluctuation theorem, which holds in the long time limit (1.78),

$$\langle \dot{S} \rangle = k \lim_{\tau\to\infty} \frac{1}{\tau} \left\langle \ln \frac{\rho_\tau(S)}{\rho_\tau(-S)} \right\rangle, \tag{2.28}$$

$$= k \lim_{\tau\to\infty} \frac{1}{\tau} D[\rho_\tau(S) || \rho_\tau(-S)]. \tag{2.29}$$

where $p_\tau(S)$ is the probability to observe an entropy production S in the interval $[0, \tau]$. Notice that the average in (2.28) is done over all possible values of the entropy production by averaging with $\rho_\tau(S)$, which yields the KLD in (2.29). Comparing (2.27) and (2.29) we arrive at

$$D\left[\mathcal{P}\left(\{z(t)\}_{t=0}^{\tau}\right) || \mathcal{P}\left(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau}\right)\right] = D[\rho_\tau(S) || \rho_\tau(-S)], \tag{2.30}$$

for $\tau \to \infty$. Consequently, although S is another observable that is obtained as a function of the micro state of the system, the KLD calculated with S yields the same value as the one calculated with full information of the system. Therefore entropy production captures all the information about the time irreversibility of the NESS.

When one does not observe the entire microscopic trajectory $\{z(t)\}_{t=0}^{\tau}$ in (2.27) but the trajectory followed by one or several observables of the system $x(t)$, the KLD only provides a lower bound to the entropy production, as we discussed in Chapter 2.2. Equation (2.30) indicates that the equality is recovered if the observables determine in a unique way the entropy production or the dissipated work.

We are interested in exploring this formula in simulations and experiments in which a microscopic system reaches a NESS and we are given a single stationary trajectory or time series produced by the system. In an experimental context, the observables are usually sampled at a finite frequency. The output is then a time series of data or discrete trajectory, $\mathbf{x} = (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n)$, where $\hat{x}_i$ can be the value of a single or several observables of the system. In

this case, we are interested in estimating the entropy production *per data* produced by the underlying physical process, which we denote by $\langle \dot{S} \rangle$ in the rest of the chapter. Entropy production per data is related to the KLD rate *per data*, which we define below.

For the sake of simplicity, we now consider random discrete processes, but the discussion below holds also for continuous random processes. Given an infinitely long realization or time series sampled from a random discrete process X_i ($i = 1, 2, \dots$), which can be multi-dimensional, we define by $p(x_1^m)$ the probability that a given sequence of m consecutive data is equal to $x_1^m = (x_1, x_2, \cdots, x_m)$. We define the m−th order KLD for this random process X_i by the distinguishability (KLD) between $p(x_1^m)$ and the probability $p(x_m^1)$ to observe the reverse sequence of data $x_m^1 = (x_m, x_{m-1}, \cdots, x_1)$.

m−th order KLD rate

$$D_m^X = D[p(x_1^m) \| p(x_m^1)] = \sum_{x_1, \cdots, x_m} p(x_1^m) \ln \frac{p(x_1^m)}{p(x_m^1)}. \tag{2.31}$$

The *KLD rate* for the process X_i is defined as the growth rate of D_m^X with the number of data,

KLD rate

$$d^X = \lim_{m \to \infty} \frac{D_m^X}{m}. \tag{2.32}$$

Because of the finite time sampling and given that x may not contain the information of the entropy production (2.30), the chain rule (2.17) implies that the KLD rates bounds from below the entropy production per data

Irreversibility and dissipation in the NESS

$$\frac{\langle \dot{S} \rangle}{k} \geqslant d^X. \tag{2.33}$$

The above bound is saturated if the random variable is the microstate of the system $X = \{\mathbf{q}, \mathbf{p}\}$ and the sampling rate is infinite or X determines uniquely the entropy production in the process.

Equation (2.33) is our basic result. It reveals a striking connection between physics and the statistics of a time series. The left-hand side, $\langle \dot{S} \rangle / k$, is a purely physical quantity, whereas the right-hand side, d^X, is a statistical magnitude depending solely on the observed data, but not on the physical mechanism generating the data. This means that if we are given a stationary time series of any random variable X produced by a microscopic system we can bound from below the average entropy production rate in the physical mechanism that generated the data. In particular (2.33), can be used to study the minimum amount of entropy produced in a symmetry restore such as the erasure of a bit, which yields Landauer's principle relating entropy production and logical irreversibility in computing machines [Kaw07, Lan61, Ando8a] as we will see in Chapter 6. Equation (2.33) extends this principle and suggests that we can determine the average dissipation of an arbitrary NESS, even ignoring any physical detail of the system.

2.4 ANALYTICAL RESULTS FOR DISCRETE SYSTEMS

We now proceed to study the bound to the entropy production provided by the KLD in the NESS (2.33) in two specific processes in which the KLD can be calculated analytically or semi-analytically, a Markov chain satisfying detailed balance condition and a particular case of a Hidden Markov chain. The results above were first introduced in two of our works [Rol10, Rol12].

2.4.1 *Markov chains obeying local detailed balance*

We first analyze how the bound (2.33) is expressed for Markovian time series that obey detailed balance by deriving analytical expressions for both entropy production and the KLD rate.

Among all the stochastic processes, the *Markov* processes are the most important ones in physics, chemistry and biology [VK92]. Let us consider a stochastic process of a random (discrete or continuous) variable X. Such a process is said to be *Markovian* if the probability to observe any sequence of n data x_1^n at times $t_1^n = (t_1, t_2, \cdots, t_n)$ satisfies the following property,

$$\rho(x_{n+1}, t_{n+1}|x_1, t_1, \cdots, x_n, t_n) = \rho(x_{n+1}, t_{n+1}|x_n, t_n), \quad (2.34)$$

where the bar | denotes conditioned probability. Therefore, in a Markov process, the probability to observe a value of the process at a given time t_{n+1} only depends on the state of the process one step before, at time t_n. In order to know the probability to observe a sequence x_1^n, we only need to know $p(x_1, t_1)$ and the transition probability $p(x_2, t_2|x_1, t_1)$ for successive times t_1, t_2 and any value of x_1, x_2 ,since the following property derives from (2.34)

$$\rho(x_1, t_1; \cdots; x_n, t_n) = p(x_1, t_1) \cdot p(x_2, t_2|x_1, t_1) \cdots p(x_n, t_n|x_{n-1}, t_{n-1}). \quad (2.35)$$

Nuclei decay, the voltage in an RC circuit, or the motion of a molecular motor in a microtubule can be described by Markov processes. Concerning microscopic physics, the position of a Brownian particle described by the overdamped Langevin equation can be considered as a Markov process [VK92]. The probability of the random variable X to take the value x at time t, $p(x, t)$ obeys the *Master equation*

Master equation (continuous variable)

$$\frac{\partial p(x,t)}{\partial t} = \int dx' \left[k(x|x')p(x',t) - k(x'|x)p(x,t)\right], \quad (2.36)$$

where $k(x|x')$ is the *rate* or transition probability per unit of time from x' to x. If X can only take discrete values, and we define by

p_i the probability of X to be at any state i at time t, the Master equation for the probability distribution is

Master equation (discrete case)

$$\dot{p}_i = \sum_j k_{j\to i} p_j - k_{i\to j} p_i, \tag{2.37}$$

where $k_{i\to j}$ is the rate from state i to state j, *i.e.* the number of times that the transition $i \to j$ occurs per unit of time. Therefore, $k_{i\to j} \geqslant 0$ for all i, j. The master equation (2.37) can be seen as a gain-loss equation for state i. The term $k_{j\to i} p_j$ accounts for the net incoming probability from any state j to state i and the term $-k_{i\to j} p_i$ accounts for the losses from state i to any other state j. The net change of p_i due to an exchange with state j is defined as the *current*

Current between two states

$$J_{j\to i} = k_{j\to i} p_j - k_{i\to j} p_i, \tag{2.38}$$

which allows one to write the master equation as a balance equation

$$\dot{p}_i = \sum_j J_{j\to i}. \tag{2.39}$$

A Markov process reaches a stationary state when the probability to be at any state i does not change in time, $\dot{p}_i = 0$, which implies by (2.37) the balance condition $\sum_j J_{j\to i} = 0$ for all i. If the following (stronger) condition called *detailed balance* condition is satisfied for every pair of states i and j,

$$\frac{k_{i\to j}}{k_{j\to i}} = \frac{p_j}{p_i} \tag{2.40}$$

the process also reaches a stationary state. Notice however that detailed balance condition on the transition rates $k_{i\to j}$ *does not* ensure that the stationary state is an equilibrium state. If the system is in contact with a thermal bath at temperature T during the process, the following constraint called *local detailed balance* condition

Local detailed balance

$$\frac{k_{i\to j}}{k_{j\to i}} = \exp\left(-\frac{V_j - V_i}{kT}\right), \tag{2.41}$$

has to be satisfied for every i, j in order to reach a stationary equilibrium state. Equation (2.41) is compatible with stationary probabilities that are weighed by the Boltzmann factor, $p_i^{ss} = p_i^{eq} \propto \exp(-\beta V_i / kT)$. In fact, local detailed balance condition is derived by imposing equilibrium stationary probabilities in detailed balance condition (2.40).

When both the state x_1^n and the times t_1^n are discrete, the Markov process is called a *Markov chain* [VK92]. A simple example of a three-state Markov chain is shown in Fig. 12. A physical

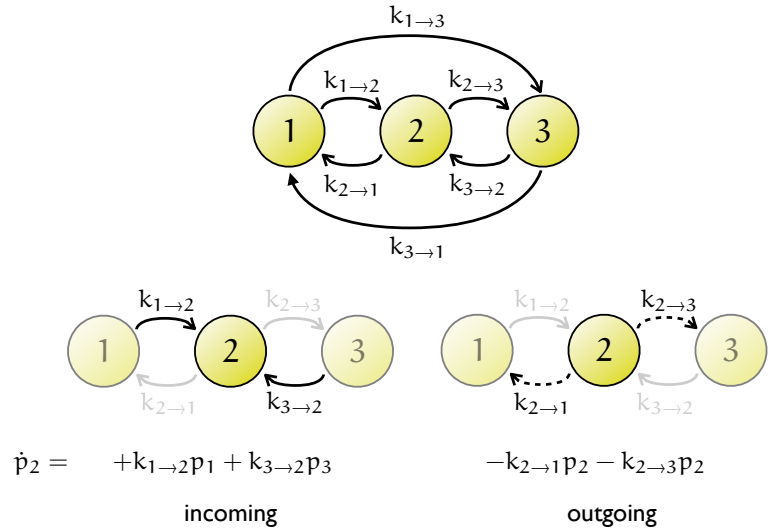


Figure 12: Example of a Markov chain. A system can jump randomly between three states labeled by 1, 2 and 3. When the system is at state i, its state in the next step is $j \neq i$. The transition rates are indicated in the top figure. In the bottom figure, the incoming and outgoing contributions for the time derivative of p_2 are illustrated (solid lines for incoming and dashed for outgoing). In the bottom line the Master equation for the probability to be in state 2 is written.

system jumps at discrete times $t = 0, \Delta t, 2\Delta t, \cdots$ between three states labeled by an index $s = 1, 2, 3$. The transition rates $k_{i\to j}$ and the balance equation for the probability to stay in one of the three states are illustrated in Fig. 12.

Let us now consider a random Markov chain X_i where the random variable can only take discrete values and we can only sample the value of the random variable at a finite frequency. For a Markov chain, the probability distribution to observe a sequence x_1^m, $p(x_1^m)$, factorizes $p(x_1^m) = p(x_1)p(x_2|x_1)\cdots p(x_m|x_{m-1})$, which also holds if we reverse the arguments, i.e., for $p(x_m^1)$. Substituting these expressions into equation (2.32), we get

KLD of a Markov chain

$$d^X = \sum_{x_1,x_2} p(x_1,x_2)\ln\frac{p(x_2|x_1)}{p(x_1|x_2)} = D_2^X - D_1^X = D_2^X, \qquad (2.42)$$

since $D_1^X = 0$ when comparing a trajectory and its reverse. Therefore, d^X only depends on transition probabilities if X is a random Markovian process. This expression was also derived for Markov chains in [Gas04].

We now relate d^X in Eq. (2.42) with the entropy production when the system reaches a NESS, because it is in contact with several thermal baths. In this situation, the local detailed balance condition is satisfied. We call $V(x_i)$ is the energy of the state x_i, and T_{x_1,x_2} is the temperature of the bath that activates the

transitions $x_1 \to x_2$ and $x_2 \to x_1$. The local detailed balance condition reads in this case

$$\frac{p(x_2|x_1)}{p(x_1|x_2)} = \exp\left(-\frac{V(x_2)-V(x_1)}{k\,T_{x_1,x_2}}\right). \tag{2.43}$$

Inserting (2.43) into (2.42),

$$\begin{aligned} d^X &= \sum_{x_1,x_2} p(x_1,x_2)\frac{V(x_1)-V(x_2)}{k\,T_{x_1,x_2}} \\ &= \sum_{x_1,x_2} p(x_1,x_2)\frac{Q_{x_1,x_2}}{k\,T_{x_1,x_2}} = \frac{\langle \dot{S}\rangle}{k}, \end{aligned} \tag{2.44}$$

where $Q_{x_1,x_2} = V(x_1) - V(x_2)$ is the heat dissipated to the corresponding thermal bath in the jump $x_1 \to x_2$, and $\dot{S}$ is the total entropy production per data. Therefore, Eq. (2.33) is reproduced, with equality, in the case of a physical system obeying local detailed balance, if we have access to all the variables describing the system. The same conclusion is reached if we induce the NESS by means of non-conservative constant forces.

Equation (2.42) can be explored further by means of the current from the state x_1 to the state x_2 as the net probability flow from x_1 to x_2, $J_{x_1\to x_2} = p(x_1,x_2) - p(x_2,x_1)$. If the system is not far from equilibrium the current tends to zero, and the following condition is satisfied $J_{x_1\to x_2} \ll p(x_1,x_2)$, yielding

$$\frac{\langle \dot{S}\rangle}{k} = d^X = D_2^X \simeq \sum_{x_1,x_2} \frac{(J_{x_1\to x_2})^2}{2p(x_1,x_2)}. \tag{2.45}$$

This expression is well known from linear irreversible thermodynamics [Sei05, Par02], where entropy production is given by the product of a flow times a thermodynamic force that is proportional to the flow itself [cf. Equation (2.1.2)]. Equation (2.45) implies that the time asymmetry of a Markovian process not far from equilibrium is revealed by the currents or probability flows that can be observed. In other words, a Markovian process without flows is time reversible. This is not the case for non-Markovian time series, where irreversibility can show up even in the absence of currents (see below and [Rol10, Rol12]).

2.4.2 *Hidden Markov processes*

In many experimental situations, a physical process is Markovian at a micro- or mesoscopic level of description, but the observed time series only contain a subset of the relevant observables, being non-Markovian in general. This is the case in biological systems, where one can only register the behavior of some mechanical and maybe a few chemical variables, while most of the relevant chemical variables cannot be monitored. These kind of

non-Markovian time series obtained from an underlying Markov process are called *Hidden Markov processes* [Rab89]. If the time and the state of the system are both discrete, the process is called a *Hidden Markov chain*. A simple example of a Hidden Markov chain is shown in Fig. 13.

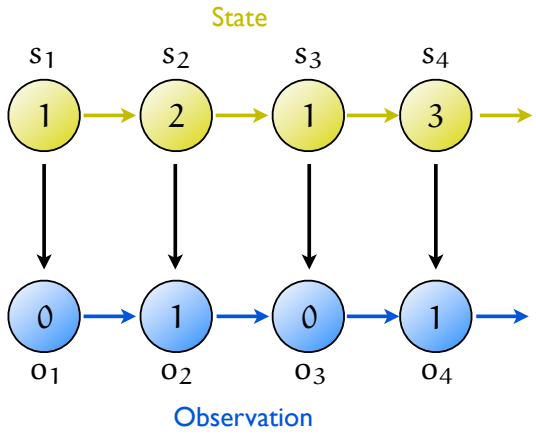


Figure 13: Example of a Hidden Markov chain. The state of a system s changes in discrete time steps according to the Markovian process described in Fig. 12. At every time, the observed state o is obtained according to the following rule: If $s_i = 1$, then $o_i = 0$, else $o_i = 1$. The sequence of observations $o_1, o_2, o_3, \cdots$ does not form a Markov chain while the (hidden) state does.

We derive a semi-analytical technique to calculate the KLD rate between hidden Markov chains. We focus on a simple case where an underlying Markov process is described by two observables X and Y; however we only observe X whose evolution is described by a hidden Markov chain. The KLD rate for the observable X is

$$d^X = \lim_{m\to\infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \frac{p(x_1^m)}{p(x_m^1)}. \tag{2.46}$$

$$= \lim_{m\to\infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \frac{\sum_{y_1^m} p(x_1^m, y_1^m)}{\sum_{y_m^1} p(x_m^1, y_m^1)}. \tag{2.47}$$

where we have expressed the marginal distribution in X by summing the joint distribution in X, Y to all the possible values that Y can take, $p(x_1^n) = \sum_{y_1^n} p(x_1^n, y_1^n)$. Because of the chain rule the KLD for the random variable X is smaller than the KLD calculated with full information given by x and Y, $d^X \leqslant d^{X,Y}$.

To compute d^X analytically, it is convenient to write d^X as a difference between two terms, $d^X = h_r^X - h^X$, where

$$h^X = -\lim_{m\to\infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln p(x_1^m), \tag{2.48}$$

$$= -\lim_{m\to\infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \sum_{y_1^m} p(x_1^m, y_1^m), \tag{2.49}$$

$$\tag{2.50}$$

is called *Shannon entropy rate*, and it is the discrete version of the Shannon entropy introduced in stochastic thermodynamics (1.66). On the other hand,

$$h_r^X = -\lim_{m\to\infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln p(x_m^1), \tag{2.51}$$

$$= -\lim_{m\to\infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \sum_{y_m^1} p(x_m^1, y_m^1), \tag{2.52}$$

$$\tag{2.53}$$

is called the *cross entropy rate*. Since the underlying process is Markovian, the probability distribution in X, Y variable factorizes $p(x_1^m, y_1^m) = p(x_1, y_1)p(x_2, y_2|x_1, y_1) \cdots p(x_m, y_m|x_{m-1}, y_{m-1})$ (2.35) and both Shannon and cross entropy can be expressed in terms of the trace of a product of random transition matrices $\mathbf{T}$ [Jaco8, Holo6]. These are square $M \times M$ random matrices, where M is the number of values that the variable y can take on, and their entries are given by

$$\mathbf{T}(x_1, x_2)_{y_1 y_2} = p(x_2, y_2|x_1, y_1). \tag{2.54}$$

There are a total number of N^2 transition matrices, where N is the number of values that x can take on. Note the different role played by each variable in this formalism: x_i are parameters defining the matrix (making $\mathbf{T}$ a random matrix), whereas y_i are subindices of the matrix elements. The Shannon and cross entropy can be expressed in terms of these matrices,

$$h^X = -\lim_{m\to\infty} \frac{1}{m} \left\langle \ln \mathrm{Tr} \left[\prod_{i=1}^{m-1} \mathbf{T}(x_i, x_{i+1}) \right] \right\rangle, \tag{2.55}$$

$$h_r^X = -\lim_{m\to\infty} \frac{1}{m} \left\langle \ln \mathrm{Tr} \left[\prod_{i=1}^{m-1} \mathbf{T}(x_{m-i+1}, x_{m-i}) \right] \right\rangle \tag{2.56}$$

where $\langle\cdot\rangle$ denotes the average over the random process X_i, which is weighted by $p(x_1^m)$. For sufficiently large m, Eqs. (2.55) and (2.56) are self-averaging [Holo6], meaning that we do not need to

calculate the average but just compute the trace for a single stationary trajectory. For any sufficiently long time series $\mathbf{x} = (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n)$ with n large, the following expressions converge to $-h$ and $-h_r$ almost surely [1]

$$\hat{\lambda}^{\mathbf{x}} = \frac{1}{n} \ln \left\| \prod_{i=1}^{n-1} \mathbf{T}(\hat{x}_i, \hat{x}_{i+1}) \right\| \simeq -h^X, \tag{2.57}$$

$$\hat{\lambda}^{\tilde{\mathbf{x}}} = \frac{1}{n} \ln \left\| \prod_{i=1}^{n-1} \mathbf{T}(\hat{x}_{n-i+1}, \hat{x}_{n-i}) \right\| \simeq -h_r^X, \tag{2.58}$$

where $\|\cdot\|$ is any matrix norm that satisfies $\|\mathbf{A}\cdot\mathbf{B}\| \leqslant \|\mathbf{A}\|\,\|\mathbf{B}\|$ [Holo6]. In particular, the trace satisfies this condition for positive matrices. In the context of random matrix theory, $\hat{\lambda}^{\mathbf{x}}$ and $\hat{\lambda}^{\tilde{\mathbf{x}}}$ are known as *maximum Lyapunov characteristic exponents* [Cri93] and measure the asymptotic rate of growth of a random vector when being multiplied by a random sequence of matrices. In practice, we can estimate d^X semi-analytically as

$$\hat{d}^{\mathbf{x}} = \hat{\lambda}^{\mathbf{x}} - \hat{\lambda}^{\tilde{\mathbf{x}}}. \tag{2.59}$$

Here $\hat{\lambda}^{\mathbf{x}}$ and $\hat{\lambda}^{\tilde{\mathbf{x}}}$ are estimated using (2.57) and (2.58) with a single time series $\mathbf{x}$ of size n, following a technique introduced in Ref. [Cri93]:

1. We generate a random stationary time series $\mathbf{x} = \{\hat{x}_1^n\}$ and compute the matrices $\mathbf{T}$ analytically.
2. A random unitary vector is multiplied by those matrices in the order given by (2.57) and normalized every l data, keeping track of the normalization factor.
3. The product of these factors divided by n yields $\hat{\lambda}^{\mathbf{x}}$.
4. For $\hat{\lambda}^{\tilde{\mathbf{x}}}$, the same procedure is repeated but using the reversed time series $\tilde{\mathbf{x}} = \{\hat{x}_n^1\}$.
5. The KLD is estimated using Eq. (2.59).

The technique is semi-analytical since the transition probabilities are known analytically but a single random stationary time series $\mathbf{x}$ is necessary to estimate d^X with the multiplication of n transition matrices that are chosen according to $\mathbf{x}$.

Let us recall that the estimator $\hat{d}^{\mathbf{x}}$ cannot be applied to empirical time series unless we know the Markov model behind the data. Consequently, it is not useful in practical situations. However, we will use it to check the performance of the estimators introduced in the following chapters, which only need a single stationary time series to estimate the KLD and do not assume any

1 A sequence of a random variable X, given by $X_1, X_2, \cdots$, is said to converge almost surely to x when the probability that $\lim_{n\to\infty} X_n = x$ is equal to 1.

knowledge of the dynamics generating these data. On the other hand, one can also get analytical approximations of Eqs. (2.55) and (2.56) by using the replica trick, in an analogous way as it has been done in Ref. [DO96]. The calculation is cumbersome and is explained in Appendix B.2. Both the semi-analytical and the replica calculations are used in Chapter 4 to check the accuracy of several empirical estimators of the KLD.

3

ESTIMATING THE KULLBACK-LEIBLER DIVERGENCE

In the previous chapter, we showed how to calculate the KLD between analytically (or semi-analytically) for series where we know in advance the dynamics of the underlying physical process. We now investigate how the KLD rate defined in (2.32) can be estimated from a single empirical stationary trajectory, obtained from a stochastic process whose dynamics is unknown. The empirical stationary trajectory contains n data of one or several random variables denoted by the letter X. In general, we are given a stationary time series that contains n data obtained from sampling the random variable X in n successive samplings,

$$\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_i, \hat{x}_{i+1}, \cdots, \hat{x}_{n-1}, \hat{x}_n, \tag{3.1}$$

where we call $\hat{x}_i$ to the value of the $i-$th data of an empirical trajectory. We denote an empirical trajectory of the random variable X of n data like (3.1) by $\mathbf{x} = \{\hat{x}_i\}_{i=1}^{n}$.

Several recipes to estimate the KLD for stationary time series of both discrete [Ziv93, Sch96, Rac04, Por07, Cai06] and continuous [Wan05b, Wan09, And07, And08b, PC08, Por07] random variables were introduced before our work. However, a full comprehensive analysis for series of both discrete and continuous random variables has not been done until now. In this chapter, we introduce a refinement of the existing methods for estimating the KLD for discrete time series and two new techniques to estimate the KLD for continuous stationary time series. Our estimators are valid in very general situations and only require the data of a single stationary trajectory. The performance of our estimation techniques is analyzed using data from both simulation and experiments in Chapters 4 and 5, where our estimators are applied to estimate the entropy production in the NESS.

This chapter is organized as follows: In Sec. 3.1 we introduce two estimators for the KLD rate for random discrete stationary processes that are a refinement of two techniques previously introduced in the literature. We extend our study to continuous time series in Sec. 3.2, where we introduce a new technique to estimate the KLD that used autoregressive models. In Sec. 3.3 we show a different approach to estimate the KLD of both continuous and discrete time series using the *horizontal visibility algorithm*.

3.1 DISCRETE CASE

For discrete stationary time series, there are two different strategies to estimate the KLD in the literature: *plug-in* estimators [Sch96, Rac04, Cai06], based on empirical counting of sequences of data, and estimators based on compression algorithms [Ziv93]. In this section, we introduce a refinement of the these two techniques that were discussed in our papers [Rol10, Rol12] and analyze their performance for a specific example in Chapter 4.

3.1.1 *Plug-in estimators*

The simplest approach to estimate the KLD rate is known as the *plug-in* method [Wan05b], which consists of an empirical estimation of the probabilities of sequences or *blocks* [Sch96] of m data, $p(x_1^m)$, appearing in Eq. (2.31) [Rac04]. The probability to observe the sequence x_1^m, $p(x_1^m)$, is estimated empirically from simply counting the number of times that x_1^m appears in a single stationary trajectory $\mathbf{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ of size n. For a time series of n data, the total number of sequences of m data contained in the series is $n - (m-1)$. The empirical probability distribution is formally defined as

$$\hat{p}^{\mathbf{x}}(x_1^m) = \frac{1}{n-(m-1)} \sum_{p=1}^{n-(m-1)} \delta_{\hat{x}_p, x_1} \cdots \delta_{\hat{x}_{p+(m-1)}, x_m}, \quad (3.2)$$

that is, the probability to find a sequence of m data *anywhere* within the sequence $\mathbf{x}$. In other words, the empirical probability (3.2) is obtained by scanning the series $\mathbf{x}$ in blocks of consecutive m data using a sliding window protocol. An estimate of D_m^X is obtained by plugging the empirical probability distribution into Eq. (2.31):

$$\hat{D}_m^{\mathbf{x}} = D[\hat{p}^{\mathbf{x}}(x_1^m) || \hat{p}^{\mathbf{x}}(x_m^1)] = \sum_{x_1, \cdots, x_m} \hat{p}^{\mathbf{x}}(x_1^m) \ln \frac{\hat{p}^{\mathbf{x}}(x_1^m)}{\hat{p}^{\mathbf{x}}(x_m^1)}. \quad (3.3)$$

Note that the probabilities in Eq. (3.3) include the superscript $\mathbf{x}$ to emphasize that they are obtained empirically from a single stationary time series $\mathbf{x}$ and therefore depend on each particular realization. The simplest way to estimate d^X would be by taking $\frac{\hat{D}_m^{\mathbf{x}}}{m}$ for m as large as possible. However, this naive approach is not efficient. The empirical probability $\hat{p}^{\mathbf{x}}(x_1^m)$ —and therefore $\hat{D}_m^{\mathbf{x}}$— is less accurate as m increases, because the number of possible substrings x_1^m increases exponentially and the statistics shortly becomes poor. It is convenient to find alternative expressions with a fast convergence. It turns out that the slope of $\hat{D}_m^{\mathbf{x}}$ as a function of m,

$$\hat{d}_m^{\mathbf{x}} = \hat{D}_m^{\mathbf{x}} - \hat{D}_{m-1}^{\mathbf{x}}, \quad (3.4)$$

also converges to the KLD rate but faster than $\frac{\hat{D}^x_m}{m}$. Our plug-in estimator is constructed as the limit

Plug-in estimator of the KLD

$$\hat{d}^x = \lim_{m\to\infty} \hat{d}^x_m. \tag{3.5}$$

For a Markovian time series, as shown in Eq. (2.42), the limit is reached for $m = 2$, and using distributions of three or more data we only get redundant information,

$$\hat{d}^x = \hat{d}^x_2 = \hat{d}^x_m, \tag{3.6}$$

for any $m > 2$. Therefore, $\hat{d}^x = \hat{d}^x_2$ is an excellent estimator of the KLD, d^X. If $\mathbf{x}$ is a k-*th order Markov chain* (i.e., it is Markovian when considering blocks of k data $\{\hat{x}^k_1\}$), then the limit is reached for $m = k$ [Rac04], i.e.,

$$\hat{d}^x = \hat{d}^x_k = \hat{d}^x_{k+1} = \hat{d}^x_{k+2} = \cdots \tag{3.7}$$

In this case, the KLD of blocks of $m < k$ data does not capture the full irreversibility of the process, $\hat{d}^x_1 \leqslant \hat{d}^x_2 \leqslant \cdots \leqslant \hat{d}^x_k = \hat{d}^x$. The convergence of (3.5) is then expected to be fast if a time series can be approximated by a k-th order Markov chain.

If the trajectory $\mathbf{x}$ is sampled from a general non-Markovian process, one needs further information to extrapolate $\hat{d}^x_m$ for $m \to \infty$, specially when only moderate values of m can be reached. In the examples discussed below, we have found that convergence is well described by the following ansatz, proposed by Schürmann and Grassberger [Sch96] to estimate Shannon entropy rate

$$\hat{d}^x_m \simeq \hat{d}^x_\infty - c\frac{\ln m}{m^\gamma}. \tag{3.8}$$

Here c and γ are parameters that, together with $\hat{d}^x_\infty$, can be obtained by fitting the empirical values of $\hat{d}^x_m$ as a function of m. The fitting parameter $\hat{d}^x_\infty$ gives an estimation of the limit (3.5).

3.1.1.1 *Insufficient statistics*

This estimation method is efficient as long as there is sufficient statistics in the data, that is, if for every series x^m_1 that occurs in the trajectory, its reverse x^1_m is observed at least once. On the other hand, if we find empirically $\hat{p}^x(x^m_1) \neq 0$ while $\hat{p}^x(x^1_m) = 0$ for at least one case, the argument of the logarithm in Eq. (3.3) diverges, yielding $\hat{d}^x_m = \infty$. We can avoid this divergence by restricting the sum in $\hat{D}^x_m$ to sequences x^m_1 whose reverse x^1_m occur in the time series:

$$\hat{D}^x_m \to \hat{D}^{x\star}_m = \sum_{(x^m_1)^*} \hat{p}^x(x^m_1) \ln \frac{\hat{p}^x(x^m_1)}{\hat{p}^x(x^1_m)}, \tag{3.9}$$

where $(x_1^m)^* = \{x_1^m \mid \hat{p}^{\mathbf{x}}(x_1^m) \neq 0 \text{ and } \hat{p}^{\mathbf{x}}(x_m^1) \neq 0\}$. The KLD calculated with this restriction is a lower bound to its real value, $\hat{D}_m^{\mathbf{x}\star} < \hat{D}_m^{\mathbf{x}}$.

A different strategy to avoid the divergence is to artificially bias the empirical probabilities such that all of them become positive. Instead of the observed empirical frequencies, we can use the following biased frequencies [Cai06]

$$\hat{p}^{\mathbf{x}}(x_1^m) = \frac{n^{\mathbf{x}}(x_1^m) + \gamma}{\sum_{x_1^m}[n^{\mathbf{x}}(x_1^m) + \gamma]}. \tag{3.10}$$

Here $n^{\mathbf{x}}(x_1^m)$ is the number of observations of x_1^m in $\mathbf{x}$ and γ is the bias, which is a small number that prevents any of the probabilities to be zero, assigning a probability of order γ/n to sequences that are not observed. The denominator in Eq. (3.10) ensures normalization of $\hat{p}^{\mathbf{x}}(x_1^m)$.

3.1.2 *Compression-based estimator*

Ziv and Merhav introduced in Ref. [Ziv93] an estimator of the KLD rate between two probability distributions based on compression algorithms. It consists on slicing or *parsing* stationary discrete time series into smaller parts according to a specific algorithm. The slicing produces a sequence of numbers (often called a *dictionary*) that contains the same data than the original series, but it is divided into subsequences, called *phrases*. These algorithms are called *compression algorithms* because a series of n discrete numbers is always parsed in $m < n$ phrases, therefore the description of the series has been compressed to fewer data.

The estimator by Ziv and Merhav [Ziv93] is defined in terms of two concepts which are now described, the compression length of a sequence and the cross parsing length between two different sequences. Given a series $\mathbf{x} = x_1^n$, its *compression length* $c(x_1^n)$ is defined as the number of distinct phrases in which it is parsed using the Lempel-Ziv (LZ) algorithm [Ziv78]. The LZ algorithm parses a series sequentially, such that each phrase that is added to the dictionary is the shortest distinct phrase that is not already in the dictionary. For example, let us consider the series $\mathbf{x} = x_1^{11} = (0,1,1,1,1,0,0,0,1,1,0)$. The LZ sequential parsing for this example is as follows: We start with an empty dictionary Dict $= \{\,\}$ and we first store the first element of the series $x_1 = 0$ in the dictionary since it is empty, hence Dict $= \{0\}$. Then we read the next number, $x_2 = 1$, which is not already in the dictionary, so x_2 is added to the dictionary, Dict $= \{0|1\}$. The next number in x_1^{11} is $x_3 = 1$, which is already in the dictionary. Then we append to x_3 the next number of the sequence, $x_3^4 = (1,1)$. This phrase is not in the dictionary and therefore it is parsed, Dict $= \{0|1|(1,1)\}$. By doing this for all the series x_1^{11}, we obtain the following dictionary

of phrases Dict $= \{0|1|(1,1)|(1,0)|(0,0)|(1,1,0)\}$. The compression length is the number of phrases that the dictionary contains once the series $\mathbf{x}$ is completely parsed, $c(x_1^{11}) = 6$ in this example. We illustrate this simple example in Fig. 14 where we also show how the dictionary can also be built using a binary tree data structure [Bay72].

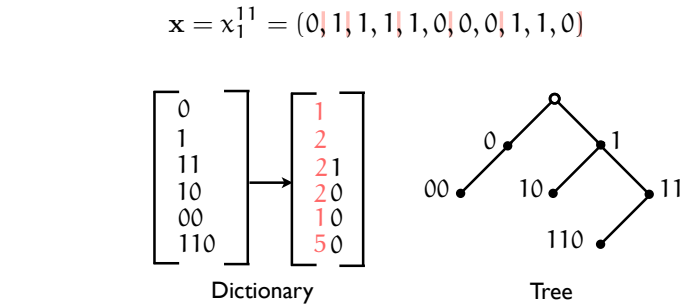


Figure 14: Lempel-Ziv compression algorithm and parsing length. We show how the LZ sequential parsing algorithm is applied to the sequence $\mathbf{x} = x_1^{11} = (0,1,1,1,1,0,0,0,1,1,0)$. The bottom left table is the dictionary of phrases obtained with the LZ algorithm. Every phrase can be encoded with two numbers: (i) A prefix number that indicates in which position on the table is the phrase that equals all the elements of the new word except the last number and (ii) a new number containing the last digit of the new word. Using this procedure, the original sequence can be described by fewer numbers than the number of data of the original sequence. The dictionary can also be built as a binary tree as shown in bottom right figure. From the parent node, new phrases are added to the tree by adding a node depending on the last value of the new word. A node to the left is added if the last number of the word is 0, and a node to the right otherwise. The compression length of the sequence is the number of phrases in the dictionary, or equivalently, the number of nodes in the tree.

The compression length of a stationary time series is related to its Shannon entropy rate [Cov06] in the limit of infinitely long sequences:

$$\lim_{n\to\infty} \frac{c(x_1^n)\ln c(x_1^n)}{n} = h^X. \tag{3.11}$$

However, since $d^X = h_r^X - h^X$, we also require an estimator for h_r^X in order to determine d^X. This is given in terms of another quantity called *cross parsing length*. The cross parsing of a series x_1^n with respect to another sequence z_1^n is obtained by parsing x_1^n looking for the longest phrase that appears anywhere in z_1^n. As an example, let us consider the cross parsing of $\mathbf{x} = x_1^{11} = (0,1,1,1,1,0,0,0,1,1,0)$ with respect to another

sequence $\mathbf{z} = z_1^{11} = (1,0,0,1,0,1,0,0,1,1,0)$. The first number in $\mathbf{x}$ is $x_1 = 0$, which is in $\mathbf{z}$. Therefore we append to x_1 the next number in $\mathbf{x}$, $x_1^2 = (0,1)$. This sequence is also somewhere in $\mathbf{z}$, more precisely it is equal to z_3^4 ,z_5^6 and z_8^9, so we append the next item in $\mathbf{x}$, $x_1^3 = (0,1,1)$. Again this sequence is somewhere in $\mathbf{z}$, $x_1^3 = z_8^{10}$, and it is added to the dictionary, $\text{Dict} = \{(0,1,1)\}$ because x_1^4 is not found anywhere in z_1^{11}. We repeat this procedure again starting from x_4 and the resulting dictionary is: $\text{Dict} = \{(0,1,1)|(1,1,0)|(0,0,1,1,0)\}$. The cross parsing length is the number of parsed sequences, which in this example is equal to $c_r(x_1^{11}|z_1^{11}) = 3$. The cross-parsing protocol for this example is shown in Fig. 15.

$$\mathbf{x} = x_1^{11} = (0,1,1,1,1,0,0,0,1,1,0)$$
$$\mathbf{z} = z_1^{11} = (1,0,0,1,0,1,0,0,1,1,0)$$

$$\mathbf{x} = x_1^{11} = (0,1,1,1,1,0,0,0,1,1,0)$$
$$\mathbf{z} = z_1^{11} = (1,0,0,1,0,1,0,0,1,1,0)$$

$$\mathbf{x} = x_1^{11} = (0,1,1,1,1,0,0,0,1,1,0)$$
$$\mathbf{z} = z_1^{11} = (1,0,0,1,0,1,0,0,1,1,0)$$

Figure 15: Cross parsing of two sequences . We highlight in green the phrases of the sequence $\mathbf{x}$ that are parsed according to the sequence $\mathbf{z}$.

In Ref. [Ziv93] it is proved that the following quantity tends to the KLD rate between the probability distributions that generated the sequences $\mathbf{x} = x_1^n$ and $\mathbf{z} = z_1^n$, which we call p^X and q^Z respectively,

$$\lim_{n\to\infty} \frac{1}{n}[c_r(x_1^n|z_1^n)\ln n - c(x_1^n)\ln c(x_1^n)] = d(p^X\|q^Z). \quad (3.12)$$

We can estimate d^X by using as inputs in the left-hand side of the above equation a stationary time series and its time reverse, that is, using $z_1^n = x_n^1$. The Ziv-Merhav estimator of d^X when using a time series $\mathbf{x}$ of n data is introduced as follows

$$\hat{d}_{ZM}^{\mathbf{x}} = \frac{1}{n}[c_r(x_1^n|x_n^1)\ln n - c(x_1^n)\ln c(x_1^n)], \quad (3.13)$$

which converges to d^X when $n \to \infty$, although the convergence is slow [Ziv93]. This estimator has been used as a measure of distinguishability in several fields such as authorship attribution [Cou05] or biometric identification [Cou10].

When the KLD rate between the probability distributions under consideration is small ($d^X \ll 1$), the estimation given by (3.12) can be even negative [Cou05]. In this case the estimator is said to be

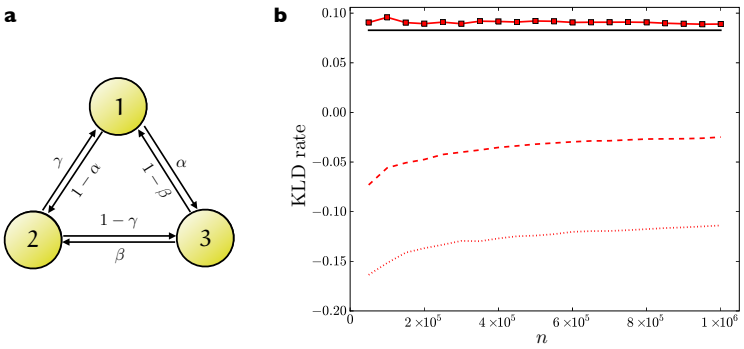


Figure 16: Validation of our KLD compression-based estimator. **a**. Sketch of the 3-state toy model used to check the accuracy of our compression estimator (3.15) and **b** comparison between different compression estimators and the analytical value of d^X. The analytical value of d^X for a model with $\alpha = 0.5, \beta = 0.7, \gamma = 0.6$ ($d^X = d_2^X = 0.08278$) is indicated by the solid black line in the plot. We show the value of the compression estimators obtained from a single stationary time series x_1^n as a function of the length n: the Ziv-Merhav estimator $\hat{d}^x_{ZM}$ (red dashed line), the bias $\tilde{d}^x_{ZM}$ (red dotted line) and our estimator $\hat{d}^x_c$ (red squares).

biased [Gra03]. The estimator gives negative values in some cases because it mixes two types of parsing: the sequential parsing of the trajectory and the cross parsing, which is not sequential. We propose the following correction, which helps to solve this issue and improves the performance of the estimator. We first evaluate (3.13) between different segments of the same trajectory. More precisely, we split $\mathbf{x}$ into two equal parts and apply the original estimator (3.12)

$$\tilde{d}^x_{ZM} = \frac{c_r(x^n_{n/2}|x_1^{n/2}) \ln \frac{n}{2} - c(x^n_{n/2}) \ln c(x^n_{n/2})}{n/2}. \tag{3.14}$$

If the time series is stationary, the two fragments, $x_1^{n/2}$ and $x^n_{n/2}$, are statistically equivalent and $\tilde{d}^x_{ZM}$ should vanish. However it is usually negative for finite n and exhibits a slow convergence to zero for large n [Cou05]. Then, we define our estimator as

Compression-based estimator of the KLD

$$\hat{d}^x_c = \hat{d}^x_{ZM} - \tilde{d}^x_{ZM}, \tag{3.15}$$

which still converges to d^X when $n \to \infty$ but it yields much better results for finite n, as we show with a simple example.

We perform a first validation of this estimator using the simplest Markov chain three-state model illustrated in Fig. 16**a**, which is equivalent to the model introduced in Chapter 2.4.1 [see Fig. 12]. Trajectories of the model are lists of numbers, 1, 2 or 3, representing the three states of the system. The dynamics is Markovian with transition probabilities given by $p_{1\to2} = 1 - p_{2\to1} = \alpha$,

$p_{2\to3} = 1 - p_{3\to2} = \beta$ and $p_{3\to1} = 1 - p_{1\to3} = \gamma$. We call X the random variable describing the state of the system and $\mathbf{x}$ a particular stationary time series, e.g. $\mathbf{x} = (1, 3, 2, 1, 2, \cdots)$. This time series is reversible only when the three transition probabilities satisfy the Kolmogorov condition [Kin69], $\alpha\beta\gamma = (1-\alpha)(1-\beta)(1-\gamma)$. In Fig. 16**b**, we compare the value of different compression estimators with the analytical value of d^X as a function of the length of the empirical trajectory n. To study how the KLD compression estimator scales with the number of data we first simulate the model in Fig. 16**a**, generating a single stationary trajectory. The estimators are measured for subsequences containing data from the first to the m−th data of the simulated series, and m is changed in the range shown in the figure. Since the trajectories described by the state of the system are Markovian, d^X only depends on transition probabilities , $d^X = d_2^X$. We see that the Ziv-Merhav estimator $\hat{d}^{\mathbf{x}}_{ZM}$ fails to estimate d^X accurately when it is small ($d^X \simeq 0.083$) and in some cases gives a negative value. The proposed estimator $\hat{d}^{\mathbf{x}}_c$, on the other hand, is significantly closer to the analytical result, although slightly overestimates its true value.

We notice that the implementation of the compression estimator is not straightforward. The construction of the LZ dictionary for a sequence requires to look-up in the dictionary every time a new phrase is added. For a series of n data, if the dictionary is constructed as a table like in the bottom left panel of Fig. 14, the search becomes slower the more phrases are stored in the dictionary. Typically, the algorithmic of this procedure is of order $O(n^2)$. To avoid this problem, Benedetto *et al* [Ben02] introduced a technique similar to the Ziv-Merhav estimator but using the length of compressed trajectories using commercial zippers instead of the LZ compression algorithm. Their approach speeds-up the estimation of the KLD but they do not provide a strong theoretical argument checking the validity of their estimator. We do not use commercial algorithms but the Ziv and Merhav compression protocol. To speed-up the estimation of the KLD, we make use of data structure trees. As shown in the bottom right panel of Fig. 14, the construction of the dictionary as a tree has the advantage that searching is faster because the information is branched. Typically, the time to search a phrase after n phrases have been added to the tree scales like $O(\log n)$, which is by far faster than using table lookups of order $O(n^2)$. We now specify step-by-step how we compute our KLD estimator using information trees:

1. The B-tree [1] (binary tree with multiple branches) of the series x_1^n is built according to the LZ sequential parsing.

1 A B-tree is the generalization of a binary tree for non-binary discrete data. Each node can have $m > 2$ children.

2. The compression length $c(x_1^n)$ is obtained as the number of nodes of the B-tree.
3. A second B-tree containing all the subsequences in x_n^1 is constructed.
4. The longest sequences in x_1^n that appear in x_n^1 are searched in the B-tree of the reverse sequence.
5. The cross parsing length $c_r(x_1^n|x_n^1)$ is obtained as the number of times in which the search in step 4 was interrupted.
6. The KLD is estimated using the Ziv-Merhav formula (3.13).

Notice that this procedure is done twice as shown in (3.15).

The main advantage of the estimation based on compression technique is that the problem with insufficient statistics arising in the plug-in estimation is avoided using this technique. The compression based estimator is also applied in Chapter 4, where its performance is compared with the plug-in technique.

3.2 CONTINUOUS CASE

For real-valued stationary trajectories, the KLD rate adopts an analogous expression to the one discussed for the discrete case in Eqs. (2.31) and (2.32). The $m-$th order KLD for a stationary process described by a continuous random variable X compares the probability densities of sequences of data $\rho(x_1^m)$ and $\rho(x_m^1)$, and is analogous to [cf. (2.31)],

$$D_m^X = D[\rho(x_1^m)||\rho(x_m^1)] = \int_{x_1^m} dx_1^m \; \rho(x_1^m) \ln \frac{\rho(x_1^m)}{\rho(x_m^1)}, \quad (3.16)$$

where $dx_1^m = dx_1 dx_2 \cdots dx_m$, and the integral above is done over the real values that $x_1 \cdots x_m$ can take. The distribution $\rho(x_m^1)$ has to be considered as a distribution $\rho_{\tilde{X}}$ for x_m^1 such that $\rho_{\tilde{X}}(x_1^m) = \rho(x_1^m)$. Therefore, the $m-$th order KLD is formally expressed as

$$D[\rho_X(x_1^m)||\rho_{\tilde{X}}(x_1^m)] = \int_{x_1^m} dx_1^m \; \rho_X(x_1^m) \ln \frac{\rho_X(x_1^m)}{\rho_{\tilde{X}}(x_1^m)}. \quad (3.17)$$

The KLD rate per data d^X is defined as in the discrete case [cf. Eq. (2.32)]

$$d^X = \lim_{m\to\infty} \frac{D_m^X}{m}. \quad (3.18)$$

We are interested in stationary time series of a continuous random variable X that is obtained by sampling X for n times at a regular

sampling rate, every Δt. In this situation, it is also interesting to study the KLD per unit of time $\dot{D}$, defined as

$$\dot{d}^X = \frac{1}{\Delta t} \lim_{m\to\infty} \frac{D_m^X}{m}. \tag{3.19}$$

Notice that the KLD rate per unit of time yields a lower bound for the average entropy rate per unit of time $\langle \dot{S} \rangle / k \geqslant \dot{d}^X$ instead of the entropy production per data.

In this section, we first review some of the first attempts in literature to estimate the KLD between random continuous processes. We also introduce a new technique that is applied to estimate the average entropy production rate in the next chapters to both simulations and experimental data.

3.2.1 *Symbolization and nearest-neighbours*

The most naive approach to measure the KLD for continuous random variables is to partition the support of the series x_1^n and x_n^1 into bins of finite size [Wan05b, Kra04]. The $m-$th order KLD for continuous variables (3.16) can be estimated by plugging in the probabilities of the series to lie within a specific bin

$$\hat{D}_m^x = \sum_{j=1}^{K} \hat{\rho}_j^x \ln \frac{\hat{\rho}_j^x}{\hat{\rho}_j^{\tilde{x}}}. \tag{3.20}$$

where $\hat{\rho}_j^x = \int_j dx_1^m \hat{\rho}^x(x_1^m)$ is the empirical probability distribution of sequences of m data integrated over the $m-$dimensional bin j. Equivalently, $\hat{\rho}_j^{\tilde{x}} = \int_j dx_1^m \hat{\rho}^x(x_m^1)$ is the same probability distribution but for the time-reversed series. By using this procedure, a continuous series is *symbolized* into a new series that contains the information of the bins of the partition visited by the series. This is a coarse-grained description of the information and therefore only a lower bound only of the actual KLD can be obtained using this method, as discussed in Chapter 2.2.

A better estimation of the KLD can be obtained by binning the continuous trajectories using data-dependent partitions [Wan05b]. The most extended technique is called the *nearest-neighbour* method, which is clarified and applied to calculate the mutual information in [Kra04]. In the nearest-neighbout method, the probabilities in (3.20) are replaced by the probabilities to find data within a $m-$th dimensional ball of radius σ centered in any data x_1^m in the $m-$dimensional space. The difference with the classical symbolization method lies in the fact that the bins in the nearest-neighbour method are centered in the data and therefore it is easier to have sufficient statistics to estimate the KLD. A comparison between the two techniques for continuous random variables was done in [Wan05b, Wan09].

More recently, another estimation of the KLD rate for continuous data has been introduced in Refs. [Ando7, Ando8b]. The concept of *pattern entropy* is defined as the Shannon entropy measured with the probability to observe sequences of a given time τ that lie within a distance σ with respect to a specific pattern or trajectory of time τ. This definition is extended to the cross entropy, and therefore to the KLD. This estimator is used in [Ando7, Ando8b] to check the relation between the dissipation and the KLD between the trajectory distributions in forward and reverse experiments (2.25) by driving a Brownian particle with an optical trap. Their result however is not extended to the stationary state and to the case in which only partial information of the system is available.

3.2.2 *Autoregressive models and KLD estimation*

We introduce a new estimator of the KLD rate for real-valued (continuous) stationary time series that is based on the use of autoregressive models. Our estimator first applies the *same* transformation $\Sigma : \mathbf{x} \to \sigma = \Sigma(\mathbf{x})$ to the time series $\mathbf{x} = x_1^n$ and the time reversal $\tilde{\mathbf{x}} = x_n^1$. The transformation yields two new time series

$$x_1^n \quad \to \quad \sigma_1^n = \Sigma(x_1^n), \tag{3.21}$$
$$x_n^1 \quad \to \quad \tilde{\sigma}_1^n = \Sigma(x_n^1). \tag{3.22}$$

We seek a transformation that satisfies the following three properties:

1. It is a one-to-one transformation in the trajectory space.
2. The new trajectories are nearly uncorrelated time series.
3. The new trajectories are not time reversal each other.

These three conditions imply, respectively: (i) The $\mathbf{x}$ and σ trajectories contain the same information and therefore the KLD rate calculated using the statistics of both yields the same value, $\hat{d}^x = \hat{d}^\sigma$, (ii) the KLD rate calculated in the new time series is approximatively the one-time KLD, $\hat{d}^\sigma = \hat{D}_1^\sigma$ and (iii) the one-time KLD in the transformed series does not vanish, $\hat{D}_1^\sigma > 0$. These three properties imply the following chain of inequalities,

$$\hat{d}^x = \hat{d}^\sigma \simeq \hat{D}_1^\sigma > 0. \tag{3.23}$$

We apply a transformation to the series that fulfills these three properties and is based on the use of autoregressive (AR) models [Box11]. We now explain in detail how do we estimate the KLD rate (3.19) from a single stationary real-valued time series.

We first fit the data x_1^n to an autoregressive model of order k and lag l, $AR(k,l)$, which is defined by a linear model of k parameters $A_1, A_2, \cdots, A_k$ such that

$$x_{1+kl} = \sum_{j=1}^{k} A_j x_{1+(k-j)l} = A_1 x_{1+(k-1)l} + \cdots + A_k x_1. \quad (3.24)$$

Given a sequence x_1^n that is sampled every l data $x_1^n \to y_1^{n/l}$, the $AR(k,l)$ model expresses the value of the new sequence $y_1^{n/l}$ at any time as a linear combination of its value at the previous k data. For lag $l = 1$, the $AR(k,l)$ model reduces to the standard $AR(k)$ model described in [Box11].

Once the parameters of the $AR(k,l)$ model are obtained from the fit, we now apply the *residual* transformation Σ to the original trajectory x_1^n. The residual function measures the deviation from the data to the $AR(k,l)$ model. If we call a_j the value of $j-$th coefficient of the fit of the data to the auto regressive model, we apply the residual function to sequences x_1^{k+1}

$$\Sigma(x_1^{(k+1)l}) \equiv x_{(k+1)l} - \sum_{j=1}^{k} a_j x_{1+(k-j)l} = x_{1+kl} - \left[a_1 x_{1+(k-1)l} + \cdots + a_k x_1\right]. \quad (3.25)$$

By applying this function to the trajectory x_1^n we get a sequence of n numbers denoted by σ_1^n. The $i-th$ component of the resulting series is obtained by applying the residual function to the substring $x_i^{i+(k+1)l}$. As an example, the first two numbers of the sequence σ_1^n are

$$\sigma_1 = \Sigma(x_1^{(k+1)l}) = x_{1+kl} - \left[a_1 x_{1+(k-1)l} + \cdots + a_k x_1\right] \quad (3.26)$$

$$\sigma_2 = \Sigma(x_2^{(k+1)l+1}) = x_{2+kl} - \left[a_1 x_{2+(k-1)l} + \cdots + a_k x_2\right] \quad (3.27)$$

We apply the same function to the time reversed series x_n^1, obtaining a different series of numbers $\tilde{\sigma}_1^n$. Once we have transformed the original series into the new ones $x_1^n \to \sigma_1^n$ and $x_n^1 \to \tilde{\sigma}_1^n$, we compute the KLD using the stationary trajectories σ_1^n and $\tilde{\sigma}_1^n$. The residual function produces a one-to-one transformation from the space of trajectories x to the space of trajectories σ such that the new time series that are obtained after the transformation are not time-reversal each other, *i.e.* $\tilde{\sigma}_1^n \neq \sigma_n^1$ and their correlation time is significantly reduced.

To do this, we first notice that if there is a one-to-one correspondence between x_1^n and σ_1^n (and x_n^1 and $\tilde{\sigma}_1^n$), the KLD rate remains the same when calculated in $\sigma-$series than when calculated for $x-$series:

$$D[\hat{\rho}^x(x_1^n) || \hat{\rho}^{\tilde{x}}(x_1^n)] = D[\hat{\rho}^\sigma(\sigma_1^n) || \hat{\rho}^{\tilde{\sigma}}(\sigma_1^n)]. \quad (3.28)$$

Here we have added a superscript $\mathbf{x}$ to stress that these are empirical probabilities of the random sequences $\mathbf{x}$. $\hat{\rho}^{\mathbf{x}}$ is the empirical probability density function of the random variable X in the forward time series. The KLD in the r.h.s. measures the distinguishability between the probability density to observe σ_1^n in the series and to observe $\tilde{\sigma}_1^n$ in the time-reversed series. The simplest approach to this number is to measure the one-time KLD rate for $\sigma-$ sequences,

$$\hat{D}_1^\sigma = \int d\sigma' \; \hat{\rho}^\sigma(\sigma') \ln \frac{\hat{\rho}^\sigma(\sigma')}{\hat{\rho}^{\tilde{\sigma}}(\sigma')}. \tag{3.29}$$

Equivalently, if the data sampling is done at a regular frequency $f_{act} = 1/\Delta t$, one can estimate the KLD rate per unit of time,

$$\dot{\hat{D}}_1^\sigma = \frac{1}{\Delta t} \int d\sigma' \; \hat{\rho}^\sigma(\sigma') \ln \frac{\hat{\rho}^\sigma(\sigma')}{\hat{\rho}^{\tilde{\sigma}}(\sigma')}. \tag{3.30}$$

These estimators contain less information than the one concerning the probability of sequences of $n > 1$ symbols, since they only take into account one-time correlations, therefore the chain rule implies that $\hat{D}_1^\sigma \leqslant \hat{d}^\sigma$. The equality is saturated only when the time series σ_1^n and $\tilde{\sigma}_1^n$ are uncorrelated, and the irreversibility is captured in the one-time statistics. In summary, using the estimators (3.29) and (3.30), we can bound from below the KLD rate (per data or per time, respectively) of the original sequences, as shown in Eq. (3.23).

In practice, we have to estimate Eqs. (3.30) and (3.30) from the data of a single stationary trajectory. We use the simplest approach which consists on partitioning the data σ_1^n and $\tilde{\sigma}_1^n$ into M bins that are equally spaced by a quantity $\Delta\sigma$ and counting the number of times that the value of the series lies within the bounds of the bin. We estimate the probability density in the $i-$th bin by $\hat{\rho}_i^\sigma = n_i/(n\Delta\sigma)$, being n_i the number of times that σ_1^n lies in bin i. Similarly, $\hat{\rho}_i^{\tilde{\sigma}} = n_i'/(n\Delta\sigma)$, being n_i' the number of times that $\tilde{\sigma}_1^n$ lies in bin i. The estimator of $\hat{D}_1^\sigma$ reads

$$\dot{\hat{D}}_1^\sigma = \frac{\hat{D}_1^\sigma}{\Delta t} = \frac{1}{\Delta t} \sum_{i=1}^{M} \Delta\sigma \; \hat{\rho_i}^\sigma \ln \frac{\hat{\rho}_i^\sigma}{\hat{\rho}_i^{\tilde{\sigma}}}. \tag{3.31}$$

We will use this formula to estimate the KLD rate of a continuous system in Chapter 5. We will denote by $\hat{d}_{AR}^{\mathbf{x}}$ the estimator of the KLD rate *per unit of time* for $\mathbf{x}-$sequences using auto regressive models, $\hat{d}_{AR}^{\mathbf{x}} = \dot{\hat{D}}_1^\sigma$, and therefore

$$\hat{d}_{AR}^{\mathbf{x}} \equiv \frac{1}{\Delta t} \sum_{i=1}^{M} \Delta\sigma \; \hat{\rho_i}^\sigma \ln \frac{\hat{\rho}_i^\sigma}{\hat{\rho}_i^{\tilde{\sigma}}}. \tag{3.32}$$

In Chapter 5 we use this estimator of the KLD rate of stationary continuous time series to distinguish between reversible and irreversible biological processes and to estimate the average energy consumption rate of the biological process in the NESS.

3.3 VISIBILITY TECHNIQUE

We now introduce a novel technique to estimate the KLD of both continuous and discrete time series that is described in our paper [Lac12]. The technique makes use of the *horizontal visibility algorithm*, which is a method that maps a time series to a directed network according to a geometric criterion. The KLD between forward and backward trajectories is estimated by the KLD between the *in* and *out* degree distributions of the associated graph.

In this section, we define the estimator based on the visibility algorithm and also show that the estimator correctly distinguishes between reversible and irreversible stationary time series of continuous random variables, including analytical and numerical studies of its performance for: (i) reversible stochastic processes (uncorrelated and Gaussian linearly correlated), (ii) reversible (conservative) and irreversible (dissipative) chaotic maps, and (iii) dissipative chaotic maps in the presence of noise.

The visibility method can also be applied to discrete data, despite in general the visibility method underperforms the plug-in method, as we show in Chapter 4, where the visibility technique is used to measure the time irreversibility of a discrete flashing ratchet in an asymmetric potential.

3.3.1 *The method*

3.3.1.1 *Complex networks*

A network is a collection of links and nodes. Any system where its different parts are coupled each other can be described by a network [Dor10]. In physics, ordered networks such as the structure of solids [Ash79] or random networks like the Ising model with random coupling matrix [She75] have been studied at first. More recently, Watts and Strogatz [Wat98] proposed a new kind of network that lies between the ordered and random network called *small-world* networks. Such a network has the features of both ordered and random networks. First, one can reach distant nodes after a few number of jumps between connected nodes, like in a random graph. Second, there are clusters within the network like in ordered graphs. During the last years, small-world networks have been found to reproduce many situations in the real world, such as metabolic networks, the world wide web, food webs [2], epidemic spreading or social ties [New03].

2 Food webs are networks that represent feeding connections in an ecological community. Nodes represent different species and two nodes are connected if one of the representing species eats the other one.

First quantitative analysis on complex networks were done in the context of random graph theory by Erdős and Renyi [Erd59, Erd61]. For the scope of this work, only two simple concepts of graph theory have to be taken into consideration. A key concept in complex networks is the *degree* of a node, k, which is defined as the number of links that connect to a specific node. The degree distribution $p(k)$ is the distribution of the degree measured over all the nodes of the network, or the probability of an arbitrary node to have degree k (i.e. k links). In some cases, links have a specific direction. For example in a food web, predators eat the prey, which can be represented by a node (predator) connected with a link that is directed to another node (prey). A network with directed links is called a directed network, and the degree at every node can be expressed as the sum between nodes that enter into the node k_{in} plus nodes that exit the node k_{out}, yielding $k = k_{in} + k_{out}$ for every node in the network. An ingoing and outgoing degree distributions, $p(k_{in})$ and $p(k_{out})$, can be defined in a network by measuring the average ingoing and outgoing degree over all the nodes in the network.

3.3.1.2 *The horizontal visibility graph*

Visibility algorithms map time series to networks according to specific geometric criteria [Lac08, Luq09]. The general purpose of such methods is to accurately map the information stored in a time series into an alternative mathematical structure. This mapping enables to employ the powerful tools of graph theory to characterize time series from a different viewpoint, bridging the gap between nonlinear time series analysis, dynamical systems, and graph theory [Lac09a, Luq11, Els09, Liu10, Yan09].

We focus here on a specific subclass of the visibility algorithms called *horizontal visibility algorithm*, firstly proposed in [Luq09]. The algorithm assigns to each data x_i of the series x_1^n to a node in the *horizontal visibility graph* (HVg). Then, two nodes i and j are connected in the graph if one can draw a *horizontal* line in the time series joining x_i and x_j that does not intersect any intermediate data height (see Fig. 17). Hence, the nodes i and j are connected if the following geometrical criterion is fulfilled within the time series:

$$x_i, x_j > x_k, \ \forall\ k \mid i < k < j\ . \tag{3.33}$$

In [Luq09, Lac10], stochastic and chaotic series are characterized using the horizontal visibility algorithm, and the first steps for a mathematically sound characterization of horizontal visibility graphs have been established in [Gut11]. The method can also be used to characterize not only time series but generic nonlinear discrete dynamical systems, sharing similarities with the theory of symbolic dynamics [Luq11].

3.3.1.3 *Directed HVg*

So far in the literature the family of visibility graphs are undirected, as visibility does not have a predefined temporal arrow. However, as already suggested in [Laco8], such a directionality can be made explicit by making use of directed networks [Newo3]. We address such directed version, defining a *Directed* Horizontal Visibility graph (DHVg) as a HVg, where the degree $k(t_i)$ of the node corresponding to the data $x_i = x(t_i)$ with $t_i = i\Delta t$, is now splitted in an *ingoing* degree $k_{in}(t_i)$, and an *outgoing* degree, such that $k(t_i) = k_{in}(t_i) + k_{out}(t_i)$. In this case, the ingoing degree $k_{in}(t_i)$ is defined as the number of links of node t_i with other *past* nodes associated with data in the series, or equivalently, the number of data that see x_i using the horizontal visibility algorithm. Conversely, the outgoing degree $k_{out}(t_i)$, is defined as the number of links with *future* nodes, or the number of data that can be seen by x_i. The construction o the DHVg associated to a time series is illustrated in Fig. 17.

We define the *in* and *out* (or ingoing and outgoing) degree distributions of a DHVg as the probability distributions of k_{out} and k_{in} of the graph which we call $p_{out}(k) \equiv p(k_{out} = k)$ and $p_{in}(k) \equiv p(k_{in} = k)$, respectively.

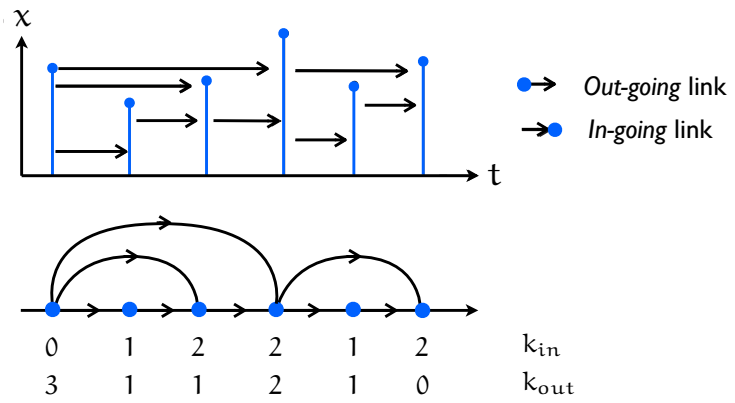


Figure 17: Graphical illustration of the directed horizontal visibility algorithm. In the top panel we plot a sample time series $x(t)$ of a continuous random variable X. A bar of height $x(t_i)$ in plotted for every data of the series. Arrows in the top panel indicate the data seen by the series at every time t_i according to the horizontal visibility criterion. The associated directed horizontal visibility graph is plotted below. In this graph, a node is associated to every data of the series. Each node has an ingoing degree k_{in}, which accounts for the number of links with *past* nodes, and an outgoing degree k_{out}, which accounts for the number of links with *future* nodes. The asymmetry of the resulting graph can be captured in a first approximation through the invariance of the outgoing (or ingoing) degree series under time reversal.

3.3.1.4 *Estimating the KLD using directed horizontal visibility graphs*

We propose a new estimator of the KLD valid for both discrete (2.32) and continuous (3.19) random variables. We claim that the time irreversibility can be captured by the KLD between the *in* and *out* degree distributions, $p_{in}(k)$ and $p_{out}(k)$. If needed, higher order measures can be used, such as the KLD between the in and out degree-degree distributions, $p_{in}(k, k')$ and $p_{out}(k, k')$. These are defined as the in and out joint degree distributions of a node and its first neighbors, describing the probability of an arbitrary node whose neighbor has degree k' to have degree k [New03].

As we will show in several examples, the information of the outgoing degree distribution k_{out} is sufficient to distinguish between reversible and irreversible stochastic stationary series which are real-valued but discrete in time x_1^n. We compare the outgoing degree distribution in the actual series $P_{k_{out}}(k|x_1^n) = p_{out}(k)$ with the corresponding probability in the time-reversed series, which is equal to the probability distribution of the ingoing degree in the actual process $P_{k_{out}}(k|x_n^1) = p_{in}(k)$. The KLD between these two distributions is

$$D[p_{out}(k)||p_{in}(k)] = \sum_k p_{out}(k) \ln \frac{p_{out}(k)}{p_{in}(k)}, \tag{3.34}$$

which vanishes if and only if the outgoing and ingoing degree probability distributions of a time series are identical, $p_{out}(k) = p_{in}(k)$, and it is positive otherwise. We define a new estimator of the KLD of a random process X from the empirical in and out degree distributions measured over the DHVg contracted from a time series $\mathbf{x}$. The empirical degree distributions associated to a time series $\mathbf{x}$ are denoted by $p^{\mathbf{x}}_{out}(k)$ and $p^{\mathbf{x}}_{out}(k)$. We define the *visibility estimator* of the KLD, $\hat{d}_{vis}$, as

Visibility estimator of the KLD

$$\hat{d}^{\mathbf{x}}_{vis} = D[p^{\mathbf{x}}_{out}(k)||p^{\mathbf{x}}_{in}(k)] = \sum_k p^{\mathbf{x}}_{out}(k) \ln \frac{p^{\mathbf{x}}_{out}(k)}{p^{\mathbf{x}}_{in}(k)}. \tag{3.35}$$

The visibility estimator has the advantage that it maps a real-valued time series into two probability distributions, $p_{in}(k)$ and $p_{out}(k)$, that have discrete support. In this case, the visibility estimation of the KLD (3.34) is much easier to calculate than the KLD for the original real-valued time series, as shown in the preceding sections. The methods to estimate the KLD for continuous data were discussed in Chapter **??** generally proceed by first making a (somewhat *ad hoc*) local symbolization of the series, coarse-graining each of the series data into a symbol (typically, an integer) from an ordered set. Then, they subsequently perform a statistical analysis of word occurrences (where a word of length n is simply a concatenation of n symbols) from the forward and

backwards symbolized series. The visibility estimator can also be considered as a symbolization since the series is mapped to a discrete series $\{k_{out}(t_i)\}_{i=1,\cdots,n}$ and $\{k_{in}(t_i)\}_{i=1,\cdots,n}$. Consequently, the estimator bounds from below the actual KLD of the time series,

$$\hat{d}^{x}_{vis} \leqslant \hat{d}^{x}. \tag{3.36}$$

However, at odds with other methods, here the symbolization process (i) lacks *ad hoc* parameters (such as number of symbols in the set or partition definition), and (ii) it takes into account *global* information: each coarse-graining $x_i \to (k_{in}(t_i), k_{out}(t_i))$ is performed using information from the whole series, according to the mapping criterion (3.33). Hence, this symbolization naturally takes into account multiple scales, which is desirable if we want to tackle complex signals [Coso5, Coso8].

3.3.2 *Validation of the method*

We now explore, analytically and numerically, the efficiency of the visibility estimator of the KLD to distinguish between reversible and irreversible stochastic stationary time series. We first analyze reversible time series generated from linear stochastic processes, which yield $\hat{d}^{x}_{vis} = 0$. This analysis is extended to chaotic signals, where our method distinguishes between dissipative ($\hat{d}^{x}_{vis} > 0$) and conservative ($\hat{d}^{x}_{vis} = 0$) chaos, and we show the robustness of the KLD of chaotic signals polluted with noise. In Chapter 4 the method is applied to a physical system in which the time irreversibility can be controlled by tuning the value of a parameter of the system. For that system, we find that the method can, not only distinguish, but also quantify the degree of irreversibility.

3.3.2.1 *Random processes*

For simplicity, we first study random uncorrelated processes. A time series of such a process is an *i.i.d.* (independent and identically distributed) sequence of numbers. As a case study, we generate a series of real-valued random $n = 10^6$ numbers distributed according to a uniform distribution that ranges from 0 to 1, $U[0, 1]$. The first data of the series are shown in the top panel of Fig. 18. The *in* and *out* degree distributions associated to the series of $n = 10^6$ data are shown in the bottom panel of Fig. 18. The distributions are, besides finite size effects, equivalent. The deviation is measured through their KLD (see table 1), yielding $\hat{d}^{x}_{vis} \sim 10^{-6}$. We will show that this number is small enough to be considered as a computational zero, $\hat{d}^{x}_{vis} \simeq 0$, which indicates the reversible character of the series.

To support our claims, we analytically show that the *in* and *out* degree distributions for *any* uncorrelated random process are

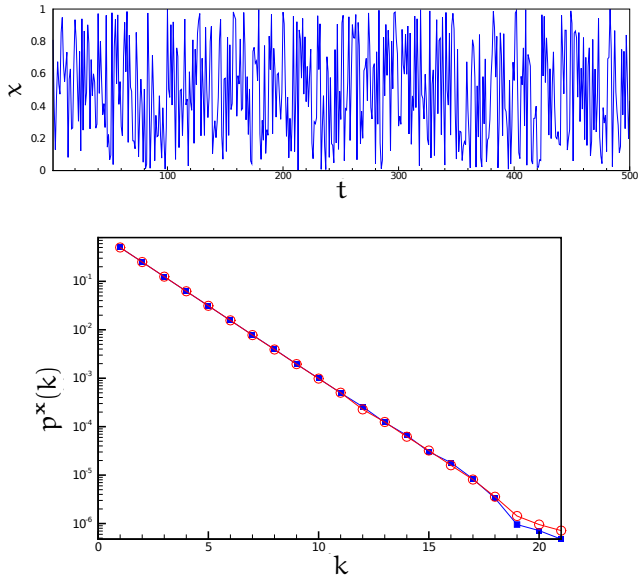


Figure 18: Ingoing and outgoing degree distributions for a random uncorrelated time series. We generate a series of $n = 10^6$ data according to a uniform distribution $U[0, 1]$. *Top*: A sample of 500 data of the series. *Bottom*: The *in* (blue filled squares) and *out* (red open circles) empirical degree distributions of the DHVg associated to the random series of 10^6 data points.

indeed identical in the limit of infinite size series. In particular, we show that the *in* and *out* degree distributions of the DHVg of an infinite sequence of i.i.d random variables extracted from a continuous probability density $\rho(x)$ are

$$P_{in}(k) = P_{out}(k) = \left(\frac{1}{2}\right)^k, \; k = 1, 2, 3, \ldots \tag{3.37}$$

The proof of this result is shown in Appendix C.1. We notice that the *in* and *out* degree distributions are identical, $d_{vis} = 0$. The random uncorrelated series are reversible at the level of description of the *in* and *out* degree distributions of the DHVg. The result is valid for any underlying probability density $\rho(x)$, it holds not only for Gaussian or uniformly distributed random series, but for any series of an i.i.d. random process.

As a further validation, we also study linearly correlated stochastic processes as additional examples of reversible dynamics [Wei75]. We use the *minimal subtraction procedure* described in Ref. [Lac10] to generate such correlated series. This method is a modification of the standard Fourier filtering method, which consists in filtering a series of uncorrelated random numbers in Fourier space.

We study both short and long-time correlated random correlated series. We generate time series of $n = 10^6$ data with an exponentially-decaying autocorrelation $C(j) = \langle x_i x_{i+j} \rangle \sim$

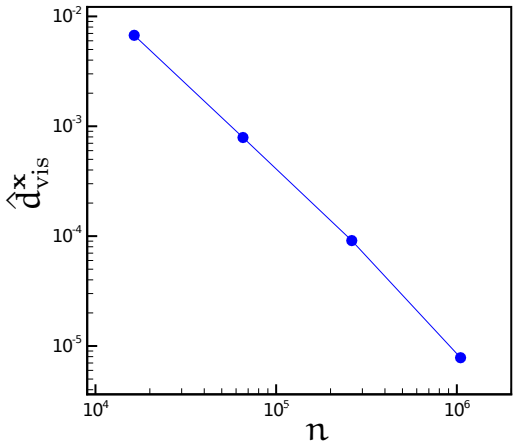


Figure 19: Scaling of the visibility estimator of the KLD for an Ornstein-Uhlenbeck process. Log-log plot of $\hat{d}^{x}_{vis}$ as a function of the number of data n of a subsequence obtained by sampling n data of a time series of 10^6 data. The original time series is generated according to an Ornstein-Uhlenbeck process of correlation $C(t) = \langle x_i x_{i+j} \rangle \sim \exp(-j/\tau)$, with $\tau = 1.0$. The blue dots are the result of an ensemble average over several realizations.

$\exp(-j/\tau)$ (akin to an *Ornstein-Uhlenbeck process*) and also using a power law decaying autocorrelation $C(j) \sim j^{-\gamma}$. In our numerical examples, we set $\tau = 1.0$ and $\gamma = 2.0$ in data units. In table 1 we show that $\hat{d}^{x}_{vis}$ is very close to zero for these series. The deviation of the KLD estimator from zero is studied in Fig. 19. We sample series of size $n' < n$ from an original series of $n = 10^6$ data and measure the KLD estimator for every series of size n'. In Fig. 19 we plot the value of $\hat{d}_{vis}$ as a function of the number of data of the subsequence, showing that the KLD estimator scales with the number of data, n, like $1/n$, and therefore it tends to 0 when n is large. Consequently, we also predict reversibility ($d_{vis} = 0$) for both short (Ornstein-Uhlenbeck) and long-time (power lay decaying) correlated random series.

3.3.2.2 *Chaos*

Chaos is defined as aperiodic long-term behavior in a deterministic system that exhibits *sensitive dependence* on initial conditions [Str01]. We remark that chaotic series are not random but deterministic, and they do not converge to stable points or periodic orbits. Sensitive dependence means that the separation between two trajectories with similar initial conditions increases exponentially with time. *Conservative* chaotic systems do conserve the distance between two trajectories whereas *dissipative* chaotic systems do not. We now apply the visibility estimator of the KLD to examples of both conservative and dissipative chaotic maps

found in [Spro1], showing that the KLD distinguishes between these two types of chaotic systems.

We first study the Logistic map defined by

$$x_{i+1} = \mu x_i(1 - x_i), \tag{3.38}$$

which models the logistic equation for population growth [Stro1]. This model exhibits chaotic behavior for $\mu \geqslant \mu_\infty = 3.569946\ldots$. We focus on the Logistic map at $\mu = 4$, which produces chaotic time series as shown in the top panel of Fig. 20, where a sample of 500 data of a series of $n = 10^6$ data is shown. The *in* and *out* degree distributions of the associated DHVg of the series are clearly distinct as shown in the bottom left panel of Fig. 20 [cf. Fig. 18]. In the bottom right panel of Fig. 20 we show the finite size scaling of the estimator $\hat{d}^x_{vis}$, which converges to an asymptotic value that clearly deviates from zero, $\hat{d}^x_{vis} = 0.377$ (see Table 1). We also prove analytically in Appendix C.2 that the *in* and *out* degree distributions are different for the logistic map and therefore $d_{vis} > 0$.

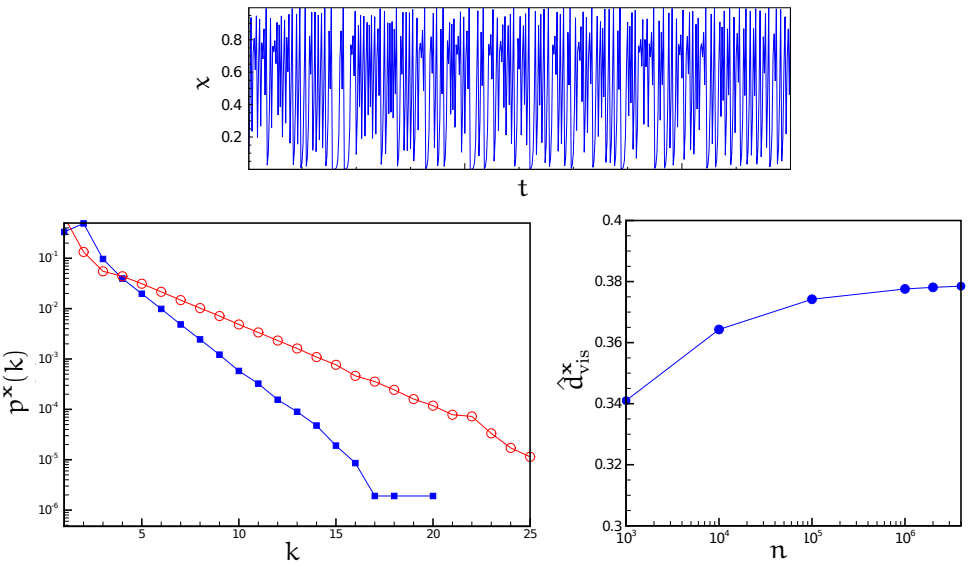


Figure 20: KLD visibility estimator for a logistic map in the chaotic region. *Top*: A sample of 500 data of a chaotic time series of $n = 10^6$ data extracted from a fully chaotic Logistic map $x_{t+1} = 4x_t(1 - x_t)$. *Bottom, left*: The *in* (blue filled squares) and *out* (red open circles) empirical degree distributions of the DHVg associated to the series. *Bottom, right*: Log-log plot of $\hat{d}^x_{vis}$ of the graph associated to a fully chaotic Logistic map $x_{t+1} = 4x_t(1 - x_t)$, as a function of the series size n (dots are the result of an ensemble average over different realizations).

For completeness, we consider other examples of dissipative chaotic systems analyzed in [Spro1]:

1. *The α-map*: $x_{t+1} = 1 - |2x_t - 1|^\alpha$, which reduces to the Logistic and tent maps in their fully chaotic region for $\alpha = 2$ and $\alpha = 1$ respectively. We analyze this map for $\alpha = 3, 4$.

2. *The 2D Hénon map*: $x_{t+1} = 1 + y_t - ax_t^2$, $y_{t+1} = bx_t$, in the fully chaotic region ($a = 1.4$, $b = 0.3$).

3. *The Lozi map*: a piecewise-linear variant of the Hénon map given by $x_{t+1} = 1 + y_n - a|x_t|$, $y_{t+1} = bx_t$ in the chaotic regime ($a = 1.7$ and $b = 0.5$).

4. *The Kaplan-Yorke map*: $x_{t+1} = 2x_t \mod(1), y_{t+1} = \lambda y_t + \cos(4\pi x_t) \mod(1)$.

We generate stationary time series using these maps and sample data once the system is in the corresponding attractor. In the case of the 2D Hénon map, the Lozi map and the Kaplan-Yorke map, we only sample the value of one of the two variables, x. In table 1 we show the value of $\hat{d}^x_{vis}$ for these series. In every case, we find an asymptotic positive value, in agreement with the conjecture that dissipative chaos is indeed time irreversible.

Finally, we also consider the *Arnold cat map*: $x_{t+1} = x_t + y_t \mod(1), y_{t+1} = x_t + 2y_t \mod(1)$. Unlike the previously discussed maps, this is an example of a *conservative* (measure-preserving) chaotic system with integer Kaplan-Yorke dimension [Spro1] [3]. We study the time irreversibility of the Arnold cat map by measuring the visibility estimator of the KLD from the data of a series of $n = 10^6$ data obtained using this map. In the top panel of Fig. 21 we show a sample of 500 data of the series generated by the Arnold cat map. The *in* and *out* degree distributions of the DHVg are indistinguishable as shown in the bottom left panel of Fig. 21. In the bottom right panel of Fig. 21 we show that d^x_{vis} asymptotically tends to zero with series size (see table 1). This correctly suggests that albeit chaotic, the map is statistically time reversible at the level of description provided by the DHVg.

In summary, our numerical study shows that dissipative chaos is time irreversible, since the KLD visibility estimator tends to a positive value $d^x_{vis} > 0$ for all the examples shown in Table 1. On the other hand, the statistics of the DHVg of the Arnold cat map suggest that conservative chaos is time reversible $d^x_{vis} \to 0$.

3 The map has two Lyapunov exponents which coincide in magnitude. $\lambda_1 = \ln(3+\sqrt{5})/2 = 0.9624$ and $\lambda_2 = \ln(3-\sqrt{5})/2 = -0.9624$. This implies that the amount of information created in the forward process (λ_1) is equal to the amount of information created in the backwards process ($-\lambda_2$), therefore the process is time reversible

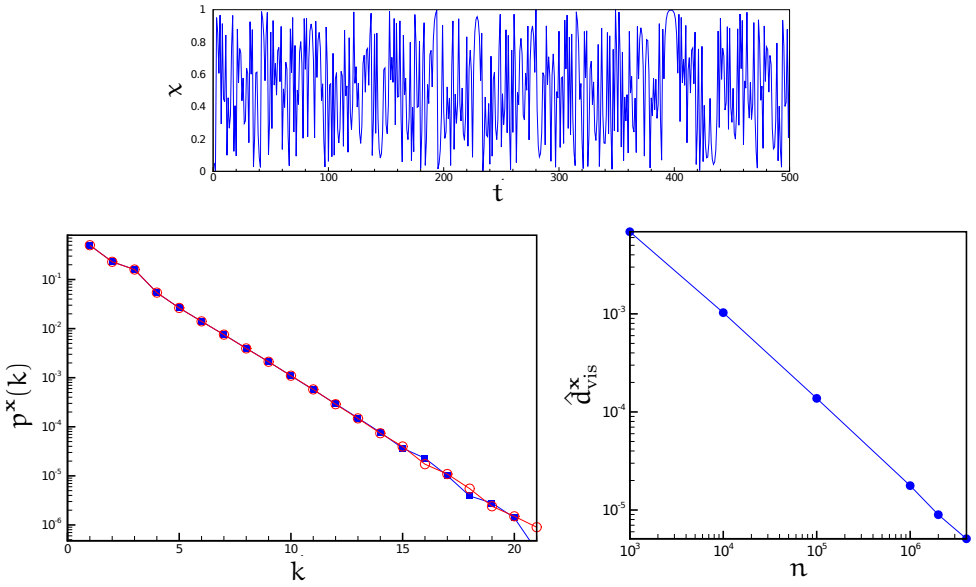


Figure 21: KLD visibility estimator for the Arnold cat map. *Top*: A sample chaotic time series (500 data points) extracted from the (chaotic and conservative) Arnold cat map. *Bottom, left*: The *in* (blue filled squares) and *out* (red open circles) degree distributions of the DHVg associated to the chaotic series of 10^6 data points. *Bottom, right*: Log-log plot of $\hat{d}^x_{vis}$ of the graph associated to the Arnold cat map as a function of the series size n. Dots are the result of an ensemble average over different realizations.

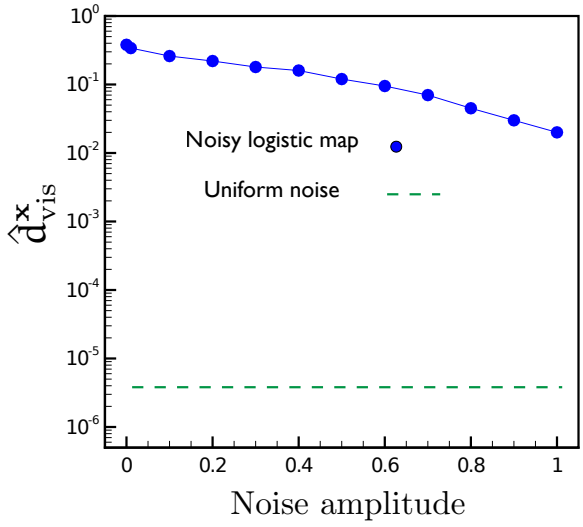
3.3.2.3 *Robustness against external noise*

Standard time series analysis methods evidence problems when noise is present in chaotic series. Even a small amount of noise can destroy the fractal structure of a chaotic attractor and mislead the calculation of chaos indicators such as Lyapunov exponents [Kos93]. In order to check if our method is robust, we add a random white noise ξ_i to a signal extracted from a fully chaotic Logistic map with $\mu = 4.0$,

$$x_{i+1} = x_i + 4x_i(1 - x_i) + \sigma\xi_i, \tag{3.39}$$

where ξ_i is a random number extracted from the uniform distribution $U[-0.5, 0.5]$ and σ is the noise intensity. We vary the amplitude of the noise σ from 0 to 1 and study how the KLD visibility estimator is robust against changes in the value of σ. In Fig. 22 we plot $\hat{d}^x_{vis}$ as a function of the amplitude of the noise. The value of $\hat{d}_{vis}$ for the noise is also plotted for comparison. The KLD of the signal polluted with noise is significantly greater than zero, since it exceeds the one associated to the noise in four orders of magnitude, even when the amplitude of the noise reaches the 100% of the signal amplitude ($\sigma = 1$). Therefore our method

correctly predicts that the signal is irreversible even when adding noise.



[h]

Figure 22: Robustness of the KLD visibility estimator against noise. Semi-log plot of $\hat{d}^{x}_{vis}$ of the graph associated to series of 10^6 data extracted from a fully chaotic Logistic map $x_{t+1} = 4x_t(1-x_t)$ polluted with white uniform noise $U[-0.5, 0.5]$ according to Eq. (3.39), as a function of the noise amplitude. The corresponding KLD value of a uniform series is plotted for comparison, which is five orders of magnitude smaller even when the chaotic signal is polluted with an amount of noise of the same amplitude.

3.3.2.4 *Table of results*

We now summarize the results obtained with the visibility estimator of the KLD in a table. In the first column of Table 1 we show the value of $\hat{d}^{x}_{vis}$ for the examples of random and chaotic processes discussed in this chapter. We also include value of the KLD obtained using the *in* and *out* degree-degree distributions, defined as

$$\hat{d}^{x}_{vis,dd} = D[p^{x}_{out}(k,k')||p^{x}_{in}(k,k')] = \sum_{k,k'} p^{x}_{out}(k,k') \ln \frac{p^{x}_{out}(k,k')}{p^{x}_{in}(k,k')}. \tag{3.40}$$

In every case we find $\hat{d}^{x}_{vis,dd} \geqslant \hat{d}^{x}_{vis}$ which is consequence of the chain rule: $D[P_{out}(k,k')||P_{in}(k,k')] \geqslant D[P_{out}(k)||P_{in}(k)]$. Note that that the method correctly distinguishes between reversible and irreversible processes, as KLD vanishes for the former and it is positive for the latter. A further analysis is done in Chapter 4 on time series produced by a physical system show that the KLD estimators bases on visibility algorithm can, not only distinguish

Series description	$\hat{d}^{x}_{vis}$	$\hat{d}^{x}_{vis,dd}$
Reversible Stochastic Processes		
$\mathcal{U}[0,1]$ uncorrelated	$3.88 \cdot 10^{-6}$	$2.85 \cdot 10^{-4}$
Ornstein-Uhlenbeck ($\tau = 1.0$)	$7.82 \cdot 10^{-6}$	$1.52 \cdot 10^{-4}$
Long-range (power law) correlated stationary process ($\gamma = 2.0$)	$1.28 \cdot 10^{-5}$	$2.0 \cdot 10^{-4}$
Dissipative Chaos		
Logistic map ($\mu = 4$)	0.377	2.978
α map ($\alpha = 3$)	0.455	3.005
α map ($\alpha = 4$)	0.522	3.518
Henon map ($a = 1.4, b = 0.3$)	0.178	1.707
Lozi map	0.114	1.265
Kaplan Yorke map	0.164	0.390
Conservative Chaos		
Arnold Cat map	$1.77 \cdot 10^{-5}$	$4.05 \cdot 10^{-4}$

Table 1: Values of the KLD visibility estimator for different stochastic processes. Empirical values of the irreversibility measure associated to the degree distribution d^{x}_{vis} and the degree-degree distribution $\hat{d}^{x}_{vis,dd}$ respectively, for the visibility graphs associated to series of 10^6 data generated from reversible and irreversible stochastic processes.

between reversible and irreversible physical dynamics, but also bound from below the entropy production in the NESS.

4

A CASE STUDY: THE FLASHING RATCHET

To explore the bound to the entropy production, we work with a discrete flashing ratchet model, where we can compare the entropy production with the analytical value and the empirical estimations of the KLD. With this model, we can analyze how information losses affect the estimation of the KLD and the tightness of the bound for the entropy production.

Flashing ratchets are particular examples of Brownian motors, which are extensively discussed in Ref. [Rei02]. Brownian motors are microscopic machines that take advantage of the thermal fluctuations of the environment and are capable to transport Brownian particles in a certain direction, and therefore to extract useful work from the fluctuating environment. The dynamics of the transported Brownian particle is irreversible and energy is dissipated to the thermal environment along the motion of the particle. Consequently, Brownian motors are excellent case studies for the relationship between irreversibility and dissipation. In the last years, Brownian motors have been used to model the dynamics of molecular biological motors [Jül97b] and artificial Brownian motors have been designed to transport colloidal particles, magnetic vortices or even charge or spin in quantum systems [Hän09].

We study one of the simplest models of Brownian motors called *flashing ratchet*, which consists of a Brownian particle moving in a linear asymmetric potential that is switched on and off randomly [Ajd92, Par04, Par02]. We use a simplified discrete version of the flashing ratchet where the dynamics of the particle is described by a Master equation between a discrete number of states. Our simple model allows one to calculate analytically the entropy production of the system and compare it to the value of the KLD rate between forward and time-reversed trajectories.

This chapter is organized as follows: In Sec. 4.1 we introduce the notion of flashing ratchet in the context of Brownian motors. In Sec. 4.2 we describe the discrete flashing ratchet model that we use in Sec. 4.3 to show that we can estimate the average entropy production of the model with the KLD of trajectories containing full information of the system. In Sec. 4.4 we show that sampling partial information of the system we can only bound from below the average entropy production of the system in the NESS, however we can still detect irreversibility even when the system does not exhibit any measurable flow.

4.1 RATCHETS

As we have already discussed, the importance of thermal fluctuations in microscopic systems gives rise to phenomena that are not observed in the macroscopic world, such as the possibility to observe negative entropy production. Another interesting effect is the possibility to extract work from thermal fluctuations. *Brownian motors* are devices that can rectify the thermal fluctuations to transport Brownian particle in a specific spatial direction. The *ratchet effect* refers to the effect of rectifying the Brownian fluctuations to produce useful work, and it was first introduced by Smoluchowski [Smo17]. The first (gedanken) experiment to observe this phenomenon was proposed by Feynman [Fey65]. Feynman's device contains an axle with vanes in one of his ends and a *ratchet* (circular saw with asymmetric saw-teeth) with pawl in the other, as we illustrate in Fig. 23. If the vanes and the ratchet are surrounded by gases at different temperatures, T_1 and T_2 respectively, the whole system moves in average in a given direction. This net movement can be used to perform work, for example a small weight can be lifted when the temperatures are different. The apparent systematic extraction of work from thermal fluctuations in Feynman's ratchet is only possible because the system is out of equilibrium. Heat flows from the hotter to the colder reservoir and the engine works irreversibly until the temperature of the two thermal baths equilibrate [Par96].

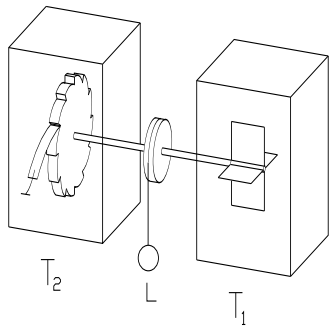


Figure 23: Feynman's ratchet and pawl motor. Picture taken from [Par96].

Since the first work by Smoluchowski, a variety of *Brownian motors* have been introduced in literature [Rei02]. These are engines that rectify thermal fluctuations producing directed transport Brownian particles. The essential ingredients of a Brownian motor are: (i) A Brownian particle immersed in a thermal bath, (ii) a spatially-periodic potential and (iii) a mechanism that drives the particle out of equilibrium. *Ratchets* are a paradigmatic exam-

ple of Brownian motors where an asymmetric potential together with a perturbation that changes the potential randomly or periodically drive the particle out of equilibrium [Ajd92]. Several different systems where directed transport is observed have been modeled as thermal ratchets. Two examples are a rotating chemical molecule [Kel99] or molecular biological motors [Jül97b]. Furthermore, artificial Brownian motors have been designed inspired in ratchets, *e.g.* using colloidal particles [Lin06], superconducting vortices [Vil03, Din07] or quantum dots [Lin98].

We focus on a specific example, namely, the *flashing ratchet*, originally described in [Ajd92]. In a flashing ratchet, a Brownian particle immersed in a thermal bath at temperature T moves in a linear asymmetric potential that is switched on and off randomly. The random switch of the asymmetric potential prevents the particle to reach thermal equilibrium and induces directed transport in a direction that depends on the shape of the potential. The simplest description for a one dimensional ratchet, is given by the Langevin equation for the position of the Brownian particle, x. The equation of motion reads

$$\gamma\dot{x}(t) = -\alpha(t)\frac{\partial V(x(t))}{\partial x} + \xi(t), \tag{4.1}$$

where $\xi(t)$ is a Gaussian white noise with zero mean and correlation $\langle\xi(t)\xi(t')\rangle = 2kT\gamma\delta(t-t')$ and $\alpha(t)$ is a dihcotomic noise that takes two possible values 0 and 1 with the same probability. When $\alpha = 0$, the particle is free and when $\alpha = 1$ the particle feels the periodic asymmetric potential described by

$$V(x) = \begin{cases} \frac{V_{max}}{a}\frac{x}{L} & \text{if } 0 \leqslant \frac{x}{L} \leqslant a, \\ -\frac{V_{max}}{1-a}\left(\frac{x}{L} - a\right) + V_{max} & \text{if } a \leqslant \frac{x}{L} \leqslant 1, \end{cases} \tag{4.2}$$

where V_{max} is the maximum height of the potential, L is the period of the potential, and a is a parameter that controls the asymmetry of the potential (only if $\alpha = 1/2$ the potential is symmetric). Eqs. (4.2), together with the periodic boundary condition $V(x \pm L) = V(x)$ describe a seesaw periodic asymmetric potential where a net particle current can be observed in the negative direction if $a > 1/2$ and in the positive direction if $a < 1/2$. In Fig. 24 we show the potential of the flashing ratchet and the probability distribution of the position of the particle when the potential is switched on and off. As it can be seen in the figure, if the potential is initially switched on, the particle first equilibrates with the thermal bath and the distribution of the position is peaked around the minimum of the potential. When the potential is switched off, the particle diffuses freely and its distribution is Gaussian. Because of the asymmetry of the potential, it is more likely that the particle jumps to the right in the potential of Fig. 24.

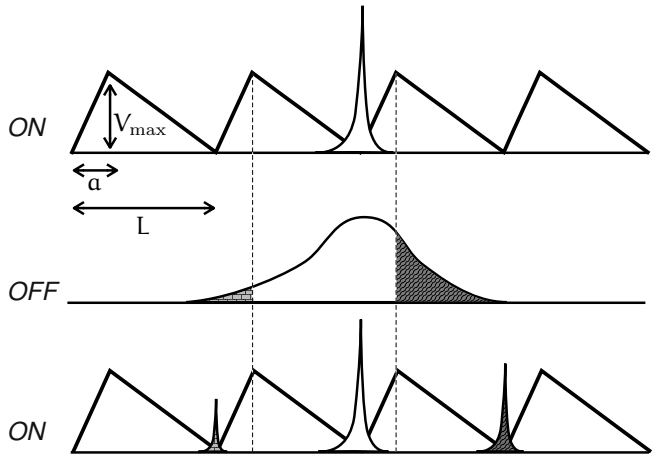


Figure 24: Flashing ratchet. Top: Seesaw asymmetric potential with parameters a, V_{max} and L indicated in the figure and probability distribution of the position of the Brownian particle shaded in white. Middle: Distribution of the position of the particle after the potential is switched off. Dark shaded areas represent regions where the particle jumps to the next or the previous period of the potential if the potential is switched on. Bottom: Equilibrium distribution of the position of the particle after the potential is switched on. As a result, a net current to the right is produced.

A net current to the right is induced in this case by switching this potential randomly or periodically [Paro2].

A discrete version of the flashing ratchet can be considered if we coarse-grain the spatial dimension in bins of equal spacing Δx. In this case, the position of the Brownian particle can be labeled by a discrete variable indicating in which bin the particle is located. By knowing the position in this coarse grained description and the state of the potential $\alpha = 0, 1$ one can describe the dynamics of the system using a Master equation for a discrete set of states as it is shown in the next section.

4.2 THE MODEL

We now explore the relation between dissipation and irreversibility in a discrete flashing ratchet model where the position of the particle is coarse grained. Our model consists of a Brownian particle moving on a one dimensional lattice. The particle is immersed in a thermal bath at temperature T and moves in a periodic, linear, asymmetric potential of height 2V, which is switched on and off at a constant rate r (see Fig. 25). Our model corresponds to the generic flashing ratchet described in the previous section with $a = 2/3$, $V_{max} = 2V$, and spatial binning of $\Delta x = L/3$. With this discretization, the state of the Brownian particle can be specified

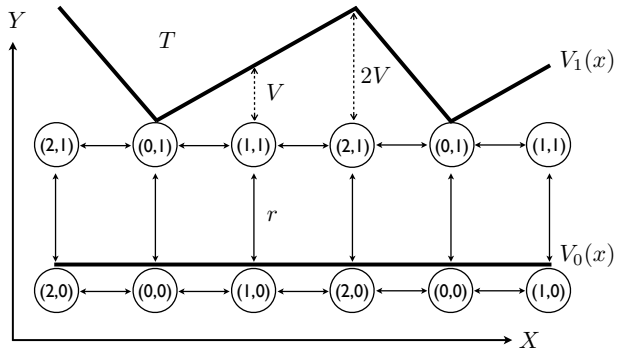


Figure 25: Illustration of our discrete ratchet model. A Brownian particle is immersed in a thermal bath at temperature T and moves in one dimension in an asymmetric linear potential $V_1(x)$ of height $2V$ and $a = L/3$ with periodic boundary conditions. The potential is switched on and off at a rate r, where $V_0(x) = 0$ represents a flat potential, and the switching probability does not depend on the position of the particle. The state of the particle is represented by two random variables (X, Y) indicated in the figure, where $X = \{0, 1, 2\}$ stands for the position of the particle whereas $Y = \{0, 1\}$ for the state of the potential. Using this description, the system can be in six different states, $(0,0), (1,0), (2,0), (0,1), (1,1)$ and $(2,1)$.

by two random observables: the position of the particle X (0, 1 or 2) and the state of the potential Y (ON, $Y = 1$ or OFF, $Y = 0$).

The particle evolves in continuous time according to a Master equation. The dynamics is described in terms of transition rates of spatial jumps and switching between six different states: $(0,0), (1,0), (2,0), (0,1), (1,1)$ and $(2,1)$. For each possible transition except switches, i.e. $(x_1, y_1) \to (x_2, y_2)$ with $y_1 = y_2 = y$, we define a transition rate $k_{(x_1,y)\to(x_2,y)}$ obeying detailed balance,

$$k_{(x_1,y)\to(x_2,y)} = \exp\left[-\frac{V_y(x_2) - V_y(x_1)}{2kT}\right]. \tag{4.3}$$

When the potential is on ($y = 1$), the value of the potential energy $V_1(x)$ is given in Fig. 25, which corresponds to (4.2) with $V_{max} = 2V$ and $a = L/3$. When the potential is off, $V_0(x) = 0$ for all x, and $k_{(x_1,0)\to(x_2,0)} = 1$ for $x_1 \neq x_2$. The switching rate does not depend on the position of the particle: $k_{(x,y_1)\to(x,y_2)} = r$ for any value of x and $y_1 \neq y_2$, and consequently violates detailed balance, driving the system out of equilibrium.

We simplify the analysis by mapping the dynamics onto a discrete-time process, a Markov chain. To this end, we record in a time series $(\mathbf{x}, \mathbf{y}) = \{x_1^n, y_1^n\}$ a list of the n first visited states, discarding any information about the time when jumps and switches occur. The resulting Markov chain is defined by the transition probabilities

$$p[(x_2, y_2)|(x_1, y_1)] = \frac{k_{(x_1,y_1)\to(x_2,y_2)}}{\sum_{x_2,y_2} k_{(x_1,y_1)\to(x_2,y_2)}}. \tag{4.4}$$

Since we discard any information about the transition times, we focus only on the dissipation and KLD rates per jump or *per data*. For finite switching rate r and since $a > 1/2$, the ratchet rectifies the thermal fluctuations inducing a current to the left in Fig. 25 [Ajd92, Par02]. The system obeys a local detailed balance condition, therefore the trajectory of the particle is a Markov chain obeying local detailed balance condition analogous to the processes discussed in Chapter 2.4.1. The nonequilibrium nature of the switching can be interpreted in two alternative ways: one can imagine that it is activated by a thermal bath at infinite temperature or by an external agent [Par02]. In either of the two interpretations, switching does not induce any entropy production (the bath needs an infinite amount of energy to change its entropy and the external agent does not produce any entropy change). Therefore, entropy is only produced when heat is dissipated to the bath at temperature T, which only occurs when the potential is on. For the time series containing the information of both position and state of the system (which we call *full* information), the average entropy production (or dissipation) per data in the time series, $\langle \dot{S} \rangle$, is then [cf. (2.44)]

$$\langle \dot{S} \rangle = \sum_{y=0,1} \sum_{x_1,x_2=0,1,2} p[(x_1,y);(x_2,y)] \frac{V_y(x_1) - V_y(x_2)}{T}. \quad (4.5)$$

Moreover, since the trajectories containing full information are Markov chains obeying detailed balance, the entropy production rate is equal to the KLD rate, $\langle \dot{S} \rangle / k = d^{X,Y} = d_2^{X,Y}$. We now analyze how we can estimate the KLD rate using single stationary trajectories of this model, and how close is this estimation to the entropy production depending on the number of degrees of freedom of the system that are sampled in the time series.

4.3 FULL INFORMATION

Firstly, we investigate the estimation of the KLD rate when using full information of the system (the position of the particle X and the state of the potential Y), and how close is this KLD rate to the actual entropy production of the process. In Fig. 26 we compare the actual dissipation and several empirical estimations of $d^{X,Y}$ for different values of the height of the potential, V. For each value of V, we simulate a single stationary time series of $n = 10^6$ data that contains full information, and calculate the plug-in estimators $\hat{d}_2^{x,y}$, $\hat{d}_3^{x,y}$, as well as the compression-based estimator of the KLD rate, $\hat{d}_c^{x,y}$. We notice that time series containing full information are constructed as follows: $(x,y) = x$ if $y = 0$ and $(x,y) = x + 3$ if $y = 1$.

Since trajectories containing full information are Markovian, the plug-in estimator immediately converges to the dissipation

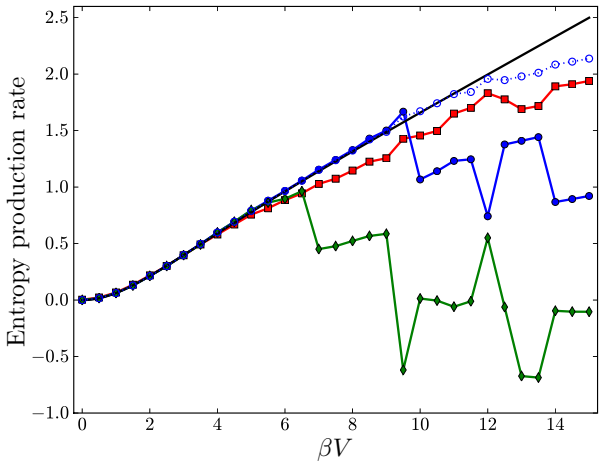


Figure 26: Analytical value of the average dissipation per data in units of kT (black line) as a function of βV in the flashing ratchet ($r = 1$) and different estimators of $d^{X,Y}$. For each value of βV, estimators are obtained from a single stationary time series of $n = 10^6$ data containing full information of the system (position, X, and state of the potential, Y): *Plug-in estimators*: $\hat{d}_2^{x,y}$ (blue circles), $\hat{d}_3^{x,y}$ (green diamonds), and $\tilde{d}_2^{x,y}$ using biased probabilities with $\gamma = 1$ (blue open circles). *Compression estimator*: $\hat{d}_c^{x,y}$ (red squares).

$\hat{d}_2^{x,y} = \hat{d}^{x,y} = d^{X,Y} = \langle \dot{S} \rangle / k$ if there is enough statistics, which happens when V is below or of order kT. If $V \gg kT$, it is very unlikely to observe uphill jumps such as $(0,1) \to (1,1)$, $(0,1) \to (2,1)$, or $(1,1) \to (2,1)$ in a single stationary trajectory. A time series of n data captures the statistics of jumps with probability well above $1/n$, which amounts to say energy jumps below $kT \ln n$, ($kT \ln 10^6 \approx 14kT$ for the trajectory used in the figures).

If, for instance, the transition $(0,1) \to (1,1)$ is missing in the trajectory, there is no way of estimating $p[(0,1);(1,1)]$ which contributes to two terms in $\hat{d}_2^{x,y}$ [see Eq. (3.3) for $m = 2$]. One of these two terms accounts for jumps $(0,1) \to (1,1)$, which are very unlikely and their contribution to the total dissipation rate is negligible, and the other term accounts for jumps $(1,1) \to (0,1)$, whose probability is larger and therefore contribute more significantly to the entropy production.

In Fig. 26, $\hat{d}_2^{x,y}$ (blue circles) and $\hat{d}_3^{x,y}$ (green diamonds) have been calculated restricting the average to sequences (of two or three data respectively) whose reverse are also observed in the time series, as given by Eq. (3.9). The sudden drops in $\hat{d}_2^{x,y}$ and $\hat{d}_3^{x,y}$ are a consequence of lack of statistics in the trajectory. For the specific time series used in Fig. 26, the lack of statistics begins at $\beta V \simeq 10$ for $\hat{d}_2^{x,y}$ and arises earlier for $\hat{d}_3^{x,y}$ because the three-data sampling space is bigger and it is easier that some transitions $(x_1, y_1) \to (x_2, y_2) \to (x_3, y_3)$ do not appear while their reverse do.

A more efficient way of dealing with the missing sequences is incorporating a small bias to the empirical probabilities, as described in Eq. (3.10). This is equivalent to assigning a probability of order $1/n$ to those transitions that are not observed in a time series of n data. Figure 26 shows the value of $\tilde{d}_2^{\mathbf{x},\mathbf{y}}$ using a biased two-data probabilities (3.10) with $\gamma = 1$ (blue open circles), which is able to extend the accuracy of the estimation even when there is lack of statistics.

Although in the case of Markovian series with a finite number of states the most convenient strategy is to use the plug-in estimator, we include for comparison the value of the compression estimator $\hat{d}_c^{\mathbf{x},\mathbf{y}}$ defined in (3.15) (red squares), which gives accurate values of the dissipation for weak potentials. Furthermore, the compression estimator is better than some plug-in estimators even for strong potentials, since it does not exhibit sudden jumps due to lack of statistics.

4.4 PARTIAL INFORMATION

We now analyze the performance of our estimators when there is not access to the full description of the system. We now assume that *only* the position of the ratchet, X, is observable. Accordingly, we simulate trajectories containing full information, and we remove the information of the state afterwards, $(\mathbf{x},\mathbf{y}) \to \mathbf{x}$. The resulting time series $\mathbf{x} = \{x_1^n\}$ are not Markovian and hence the limit (3.5) is not reached for small values of m. In this case, we proceed by obtaining $m-$data KLD rates, $\hat{d}_m^{\mathbf{x}}$, for m as large as possible and fit the resulting values to the ansatz (3.8).

We have generated trajectories of $n = 10^7$ data for values of V that range from 0 to 2kT. Once we remove the information of the state of the potential from these time series, we are able to estimate $\hat{d}_m^{\mathbf{x}}$ up to $m = 9$ with no lack of statistics. Figure 27 shows value of the plug-in estimators $\hat{d}_m^{\mathbf{x}}$ for $m = 2, 3, 5, 7, 9$ and the extrapolation $\hat{d}_\infty^{\mathbf{x}}$ (orange pentagons connected by a dashed line to guide the eye) resulting from the fit to the ansatz (3.8). For each value of βV, we fit $\hat{d}_m^{\mathbf{x}}$ as a function of m for $m = 2, 3, \cdots, 9$ to Eq. (3.8) using the curve fitting tool available in MATLAB, which provides a robust least-squares fit with bisquare weights as described in [mat]. The fit itself for a particular value of the potential, $\beta V = 1$, is shown in the inset of Fig. 27. Our ansatz reproduces the dependence of $\hat{d}_m^{\mathbf{x}}$ with m but the final estimator $\hat{d}_\infty^{\mathbf{x}}$ still bounds significantly from below the actual dissipation (black solid line in Fig. 27).

Despite plug-in estimators bound from below the entropy production, they are able to distinguish between equilibrium and NESS, even with partial information. In equilibrium ($V = 0$), the trajectories are reversible and all the estimators vanish, $\hat{d}_m^{\mathbf{x}} = 0$

for $m = 2, \cdots, 9$, whereas for the NESS ($V > 0$) they detect the irreversibility of the process yielding $\hat{d}^{x}_{m} > 0$ for all m. This is illustrated in Fig. 28, where we plot the dependence of the plug-in estimators with the size of the trajectory. For $\beta V = 0$, $\hat{d}^{x}_{2}$, $\hat{d}^{x}_{3}$ and $\hat{d}^{x}_{5}$ tend to zero when increasing the number of data whereas they saturate to a positive value in the NESS ($\beta V = 1$).

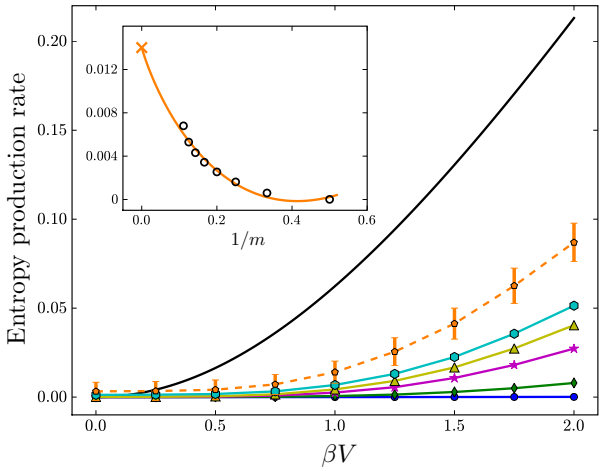


Figure 27: Average dissipation per data (black line) and plug-in estimators of d^{X} using partial information given by the position (X) for a discrete flashing ratchet with $r = 1$. For each value of βV, we calculate estimators from a single stationary time series of $n = 10^{7}$ data containing partial information: $\hat{d}^{x}_{2}$(blue circles), $\hat{d}^{x}_{3}$ (green diamonds), $\hat{d}^{x}_{5}$ (purple stars), $\hat{d}^{x}_{7}$ (yellow triangles), $\hat{d}^{x}_{9}$ (cyan hexagons) and the result from the fit $\hat{d}^{x}_{\infty}$ (orange pentagons with error bars and connected by a dashed line). *Inset*: $\hat{d}^{x}_{m}$ as a function of $1/m$ for $m = 1, \cdots, 9$ for $\beta V = 1$ (open black circles) and the fit to the ansatz (orange line). The $y-$intercept of the fit is indicated by an orange cross and it is equal to $\hat{d}^{x}_{\infty}$.

There are two possible origins for the discrepancy between $\hat{d}^{x}_{\infty}$ and the dissipation: either (i) our fit underestimates the actual KLD rate d^{X} of the trajectory; or (ii) the bound (2.33) is not tight when using only the information of the position. To address this question we need to calculate the actual value of d^{X}. Since the position of the ratchet $\mathbf{x}$ is a hidden Markov chain, we can calculate its KLD rate d^{X} semi-analytically, using the Lyapunov exponents (2.57,2.58) introduced in Chapter 2.4.2.

In Fig. 29 we show the value of the semi-analytical calculation of d^{X} using the norm of transition matrices, Eq. (2.59), which is not significantly different to the empirical estimation $\hat{d}^{x}_{\infty}$. We therefore conclude that $\hat{d}^{x}_{\infty}$ is a good estimation of d^{X}, but still d^{X} only yields a lower bound to dissipation whose accuracy is in principle hard to determine. This is an expected result, since the position of a particle in a flashing ratchet does not obey the Gallavotti-Cohen theorem [Lac09b].

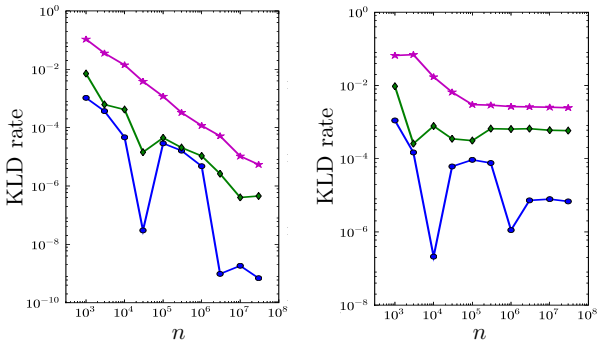


Figure 28: Scaling of plug-in estimators of d^X, $\hat{d}^x_m$, with the size of the time series n, for a flashing ratchet ($r = 1$), for $\beta V = 0$ (left) and $\beta V = 1$ (right): $\hat{d}^x_2$ (blue circles), $\hat{d}^x_3$ (green diamonds) and $\hat{d}^x_5$ (purple stars). We simulate a single stationary trajectory $\mathbf{x}$ of 10^7 data and calculate the estimators for subsequences containing the first n data of $\mathbf{x}$.

Summarizing, although $\hat{d}^x_\infty$ turns out to be a good estimator of d^X, using only information of the position we only get a lower bound to the dissipation. We also show in Fig. 29 the value of $\hat{d}^x_c$, which is well below the plug-in estimator $\hat{d}^x_\infty$. The compression estimator $\hat{d}^x_c$ lies between $\hat{d}^x_7$ and $\hat{d}^x_9$ (not shown in the plot), indicating that it is only able to capture up to 8-data correlations . For completeness, we include the calculation of d^X based on the replica trick (see Appendix B.2). The estimation using replica trick yields a tight bound for $V < kT$, but departs from d^X for larger values of V. This deviation is caused by the estimation of the limits in Eqs. (B.26,B.29), where we take $\alpha \to 0$ when α is defined only for integer values, one of the standard drawbacks of the replica trick [Cri93].

Although our estimators give low values of the dissipation when using partial information, they still capture the asymptotic behavior close to equilibrium, *i.e.* for V small. Entropy production decreases as V^2 when $V \to 0$, so do plug-in estimators $\hat{d}^x_3, \cdots, \hat{d}^x_9$, $\hat{d}^x_\infty$, and the compression estimator $\hat{d}^x_c$. Some of them are plotted in the inset of Fig. 29. On the other hand, $\hat{d}^x_2 \propto j^2 \propto V^6$, since the current is $j \propto V^3$ in this case [see Eq. (2.45)]. Recall that, close to equilibrium, calculating $\hat{d}^x_2$ is equivalent to estimating the entropy production using currents and standard linear irreversible thermodynamics, $d^x_2 \propto j^2$, as shown in Eq. (2.45). It is then remarkable that the estimators involving the statistics of three or more data are able to reproduce qualitatively the behavior of the dissipation in cases where linear thermodynamics fails.

The improvement observed when using the plug-in estimators of higher order than $\hat{d}^x_2$ is more significant in a NESS which does not exhibit observable currents in X. In this case, $j = 0$ and $\hat{d}^x_2 = 0$, but using higher order statistics we can still detect the time irreversibility of the trajectory. This happens for

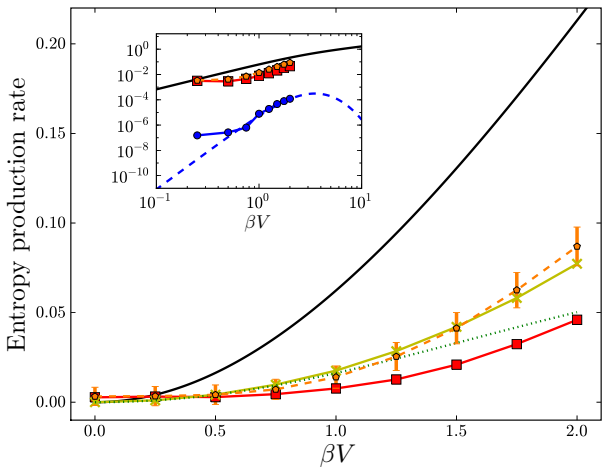


Figure 29: Average dissipation per data (black line) and different estimators of d^X for a flashing ratchet described with partial information ($r = 1$, $n = 10^7$ data) as a function of βV: $\hat{d}^x_\infty$ (orange dashed pentagons) $\hat{d}^x_c$ (red squares), replica estimation of d^X (green dotted line) and semi-analytical value of d^X (yellow crosses). *Inset*: Dependence of the average dissipation (black line), $\hat{d}^x_2$ (analytical values in blue dashed line), $\hat{d}^x_c$ and $\hat{d}^x_\infty$ on βV in the vicinity of $\beta V = 0$.

example if we add to the flashing ratchet an external force F opposite to the current, i.e., pointing in the positive x—direction. The force modifies the energy landscape and consequently the spatial transition rates $k_{(x_1 y)\to(x_2,y)}$ by a factor $e^{\beta F L_{(x_1,y);(x_2,y)}/2}$, where $L_{(x_1,y);(x_2,y)}$ is the spatial distance that separates the two points (x_1, y) and (x_2, y). Here $L_{(x_1,y);(x_2,y)}$ is defined positive if the jump $(x_1, y) \to (x_2, y)$ points in the same direction as the force (i.e. to the right), and negative otherwise. At the *stalling force* F_{stall}, the current is canceled by the force and the system does not move on average when it is described only by X, but still dissipates energy. This situation is analogous to an ATP-consuming molecular motor that stalls under the action of an external force opposite to its direction of motion [Svo94b]. At the stall force, the system looks like it is in equilibrium If we only have access to the information of the position: the spatial current vanishes, and so does $\hat{d}^x_2$, as shown in Fig. 30. However, there is a finite dissipation (black line in the figure) and the corresponding irreversibility is captured by the statistics of substrings of 3 data or more. Although d^X is below the real dissipation by an order of magnitude (see the semi-analytical value of d^X, yellow crosses in Fig. 30), it does not exhibit any sensible change at stall force. Finally, both $\hat{d}^x_\infty$ and $\hat{d}^x_c$ provide estimates of d^X which are correct within one order of magnitude (see the inset of Fig. 30).

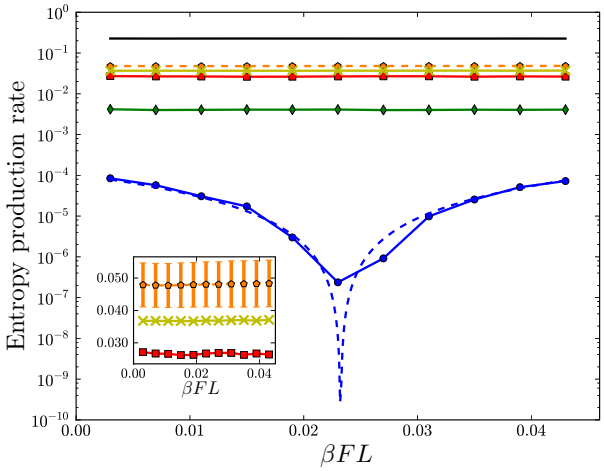


Figure 30: Average dissipation per data (in units of kT) in the flashing ratchet (with $r = 2$, and $\beta V = 2$) and different estimations of d^X obtained from a single time series of $n = 10^7$ data containing partial information (position) as a function of the external force F: analytical value of the average dissipation (black line), $\hat{d}^x_2$ (blue circles, analytical values in blue dashed line), $\hat{d}^x_3$ (green diamonds), $\hat{d}^x_c$ (red squares), semi-analytical calculation of d^X (yellow crosses) and $\hat{d}^x_\infty$ (orange hexagons). The minimum in $\hat{d}^x_2$ corresponds to the stalling force. *Inset*: $\hat{d}^x_c$, semi-analytical value of d^X and $\hat{d}^x_\infty$ as a function of the external force.

4.5 ESTIMATING DISSIPATION WITH THE VISIBILITY TECHNIQUE

We now study how our estimators of the KLD rate based on the horizontal visibility algorithm can be used to estimate the average dissipation rate in the NESS using single stationary trajectories of our discrete ratchet model. Although the visibility technique is originally designed to distinguish between equilibrium and NESS using continuous data, as shown in Chapter 3.3, we can also apply this method to discrete systems. We now go one step further and analyze if our visibility-based estimator of the KLD can also quantify the amount of irreversibility — or estimate the dissipation rate — in the discrete ratchet model. The performance of the KLD estimators using degree and degree-degree distributions are compared.

We first work with trajectories with full information generated with our discrete ratchet model. In Fig. 31 we show the values of $\hat{d}^{x,y}_{vis}$ and $\hat{d}^{x,y}_{vis,dd}$ as a function of V/kT, for stationary time series of 2^{19} data. Note that for $V = 0$, both KLD using degree distributions and degree-degree distributions vanish. On the other hand, both $\hat{d}^{x,y}_{vis}$ and $\hat{d}^{x,y}_{vis,dd}$ increase with V, providing a lower bound to the dissipation rate, which also increases with V [cf. Fig. 26]. Visibility estimators of the KLD increase with V even for large

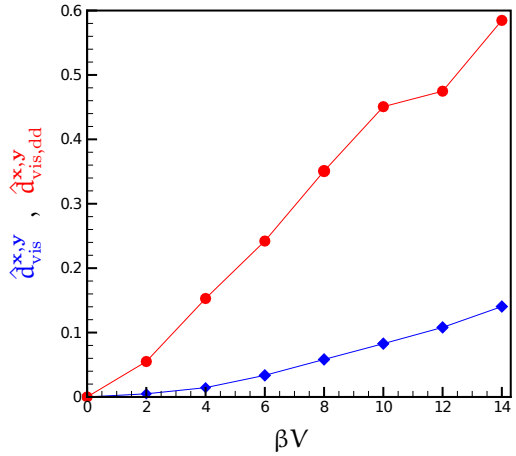


Figure 31: Visibility estimators of the KLD using full information in the flashing ratchet. $\hat{d}^{x,y}_{vis}$ and $\hat{d}^{x,y}_{vis,dd}$ for a discrete flashing ratchet ($r = 1$) as a function of V/kT. For each value of V, we generate a stationary time series of $N = 2^{19}$ steps described with full information (position and state of the potential). CHANGE Y-LABEL

values of the potential where the statistics are poor and the plug-in estimators of the KLD fail when estimating irreversibility as shown in Fig. 26. However, the value of the visibility estimators of the KLD using full information ($\hat{d}^{x,y}_{vis}$ and $\hat{d}^{x,y}_{vis,dd}$) are one order of magnitude below the dissipation and the plug-in estimators of the KLD. The degree distributions capture the irreversibility of the original series but it is difficult to establish a quantitative relationship between $\hat{d}^{x,y}_{vis}$ and the KLD between trajectories.

We now illustrate that the KLD based on the degree-degree distribution takes into account more information of the visibility graph structure than the KLD using degree distributions, providing a closer bound to the physical dissipation as it is expected by the chain rule, $\hat{d}^{x}_{vis,dd} \geqslant \hat{d}^{x}_{vis}$. The improvement is significant in some situations like the flashing ratchet with a force that opposes to the net current on the system. In Fig. 32 we address this situation, using series described with only partial information of the system (position of the particle). We show how $\hat{d}^{x}_{vis}$ tends to zero when the force approaches to the stalling value. Therefore, the value of $\hat{d}^{x}_{vis}$ is misleading in this case since it predicts reversibility, as do the KLD estimators based on local flows or currents like $\hat{d}^{x}_{2}$ (see Fig. 30). However, $\hat{d}^{x}_{vis,dd}$ captures the irreversibility of the time series, and yields a positive value at the stalling force (cf. Fig. 30).

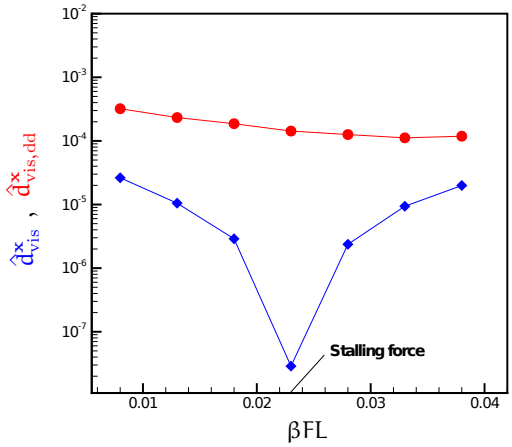


Figure 32: Irreversibility measures $\hat{d}^x_{vis}$ and $\hat{d}^x_{vis,dd}$ in the flashing ratchet ($r = 2, V = 2kT$) as a function of FL/kT. For each value of the force, we make use of a single stationary series of size $N = 10^6$ containing partial information (the state information is removed).

4.6 CONCLUSIONS

We have shown that it is possible to estimate the entropy production rate in a discrete system by analyzing statistical properties of a time series observed in a NESS. The KLD rate per data between the time series and its time reversed is a lower bound to the entropy production rate of a discrete flashing ratchet model in the NESS.

We have tested three estimators of KLD rate: plug-in estimators, the compression estimator and the estimator based on visibility algorithm. We have checked the performance of our estimators in the discrete flashing ratchet. We have shown that the KLD is a powerful tool to identify nonequilibrium states and to estimate the entropy production of a process, if this entropy production is of order of the Boltzmann constant. We have also shown that the bound given by the KLD can detect a non-zero dissipation even when the data does not exhibit any measurable flows.

Let us summarize our results by presenting a "recipe" to estimate the KLD from an experimental time series recorded from a discrete system in a NESS. If the number of possible states of the system is small enough, the best approach is to calculate the plug-in estimators $\hat{d}^x_m$ (3.4) and then check the convergence when m increases. The possible lack of statistics can be circumvented using a small artificial bias, as discussed in Sec. 3.1.1. If $\hat{d}^x_m$ saturates for some value m^*, then the time series is an m^*-th order Markov process and $\hat{d}^x = \hat{d}^x_{m^*}$. Otherwise, we can use the ansatz (3.8) and obtain $\hat{d}^x_\infty$ which is a good estimate of the KLD rate.

A second and complementary approach is the use of the compression estimator introduced in Sec. 3.1.2. The estimator yields

correct results in the examples that we have analyzed, but there is no clue about the corresponding error. Nevertheless, the compression estimator could be the only possible approach if the number of states of the time series is large. In this case, the calculation of empirical probability distributions $\hat{p}(x_1^m)$ would be unfeasible even for short substrings.

When using full information of the discrete ratchet, the estimators based on the visibility technique are sensitive to small changes in the irreversibility of the trajectories and reproduce qualitatively the dependence of the entropy production with parameters of the system. However, the bound to the dissipation is weak even using full information of the system. The second symbolization (after the first symbolization given by the discretization) produced by the visibility algorithm produces, by virtue of the chain rule, that the bound to the dissipation is weaker than using the $m-$data statistics of the trajectories. We have also found that the degree distribution captures only the information of the current whereas the degree-degree distributions can be used to predict dissipation when the current of the system vanishes.

The extension of our technique to systems with many states (or described by real-valued observables) is relevant in many practical situations, especially to analyze data coming from biological systems. This can be done considering time asymmetric functionals of the data, which reduce the number of observables, and hence the number of states, but keep information about the irreversibility of the series. In the next chapter, we address this issue by considering biological time series obtained from bullfrog's ear hair bundles.

5

APPLICATION TO BIOLOGY: THE EAR HAIR BUNDLE

Biological processes are excellent candidates to apply our technique to measure quantitatively the time irreversibility of stationary time series produced in the NESS and relate it to the entropy production rate along the process. The majority of the biological processes are nonequilibrium processes where measurable currents or flows are present. Examples of nonequilibrium biological processes are the dynamics of molecular motors, DNA packing in viruses or transport through membranes. In all these cases, the energy of the system is in part used to perform work and the rest dissipated as heat to the thermal environment. By sampling appropriate degrees of freedom, such as the position of a molecular motor along a microtubule, one can also measure time irreversibility using the KLD.

Among irreversible biological processes, *active* processes are those where the energy output is greater than the energy input and therefore cellular energy (from ATP hydrolyzation in general) is used to perform work. One example is the active ion transport in a membrane against the concentration gradient. The distinction between active and passive nonequilibrium biological process requires in principle the knowledge of many degrees of freedom of the system. In [Maro1b] however, Martin *et al.* discovered a simple method to discriminate between active and passive processes in bullfrog's ear hair bundles. By measuring only the position of the top of the hair bundle in both spontaneous and forced oscillations, they were able to distinguish between alive and dead hair cells by means of the verification of the fluctuation dissipation relationship.

In this chapter, we show how to discriminate between oscillations produced by active or passive mechanisms in ear hair bundles by using trajectories of the top of the hair bundle obtained *only* from the spontaneous oscillations. By measuring the KLD rate of trajectories of the top of the bundle we can do this distinction but also estimate the minimum energy consumption rate needed to sustain spontaneous oscillations. We apply our novel method to estimate the KLD consisting of a fit of an auto-regressive (AR) model to the original time series and the computation of the KLD between residuals obtained when that AR model is applied both to the series and to its time reverse.

We measure our KLD rate estimator using both simulated and experimental trajectories of the bullfrog's ear hair bundle. In sim-

ulations, we are able to distinguish between different qualitative behaviors and to detect bifurcations. Using experimental data extracted from bullfrog's sacculus, we are able to distinguish between active and passive oscillations and to predict that the minimum amount of dissipation is given by the ATP rate of a single adaptation motor.

This chapter is organized as follows: In Sec. 5.1 we introduce examples of irreversible processes in biology and describe the active processes, focusing on the sound transduction process done by ear hair bundles. In Sec. 5.2 we apply our technique to simulations of the ear hair bundle. Section 5.3 is devoted to distinguish between active and passive oscillations from experimental data of bullfrog's sacculus. In Sec. 5.4 we discuss the main results of this chapter.

5.1 IRREVERSIBILITY IN BIOLOGY. EAR HAIR BUNDLES.

Biological systems are among the most important examples of microscopic systems in which thermal fluctuations play an important role on the dynamics. Molecular motors, bacteria, viruses or ion channels are few examples of small biological systems that are able to perform mechanical work using the energy released from chemical reactions that occur in the environment. The majority of biological processes are driven by a chemical reaction, namely the hydrolyzation of ATP (adenosine triphospate), a chemical compound formed by a nucleoside and tree phosphate groups. ATP hydrolyzation is an exothermic (energetically favorable) chemical reaction in which the ATP looses a phosphate group and transforms into ADP (adenosine diphosphate) and an energy of order $\Delta G_{ATP} \simeq -25kT$ is released [How01][1]. The vast majority of biological processes use the energy of the ATP hydrolyzation to drive a biological system out of equilibrium, producing particle or energy flows. We now cite a few examples of biological systems depicted in Fig. 33 that are involved in such nonequilibrium processes:

- **Molecular motors**: Molecular biological machines, formed by protein subunits, can transform chemical energy into mechanical work since they are able to move in a given direction even in the presence of forces opposing the direction of the motion of the motor. *Kinesin* motors, for example, can transport vesicles by moving in a regular 8nm stepping over a protein microtubule against a maximum force of order 6pN [Svo93]. Muscle contraction is originated by the

1 Gibbs free energy ΔG is commonly used in biology since the majority of processes occur at constant temperature and pressure. A negative value of ΔG indicates that the products of a chemical reaction are less energetic than the reactives, and therefore the reaction is energetically favorable or *exothermic*.

motion of *Myosin* motors that are connected to *actin* filaments, which form the muscular tissue [Hux69]. *F1-ATPase* is a motor formed by two protein subunits that is capable to rotate in a specific direction and plays a key role in the synthesis of ATP in photosynthesis [Noj97].

- **Viruses**: Bacteriophages are viruses that can inject DNA inside a bacteria, which produces a replication cycle of the virus and the death of bacteria. It has been observed that bacteriophage $\phi 29$ can package their replicated DNA before infecting new cells. This packing is an irreversible process which can be performed even against large opposing forces of the order of magnitude of 50pN [Smi01].

- **Ion channels**: Ions can be transported through biological membranes by diffusion when the ionic concentration on the two sides of the membrane is different. Moreover, ions can also be transported *against* the concentration using the energy of ATP hydrolization, for example in the sodium-potassium pump, which is involved in the transmission of electric signals in nerves [Dun61].

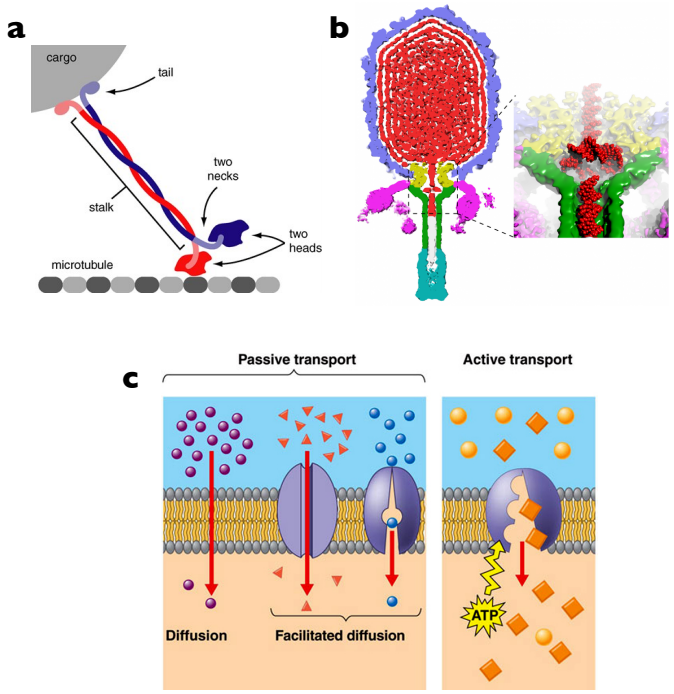


Figure 33: Nonequilibrium processes in biology. **a**. A Kinesin motor, formed by two protein subunits, can transport a cargo through along a microtubule in $8-$nm steps. Picture taken from [Kin]. **b**. Bacteriophage $\phi 29$ virus. DNA (in red) is highly packed inside the capsid of the virus. Picture taken from [Pha] **c**. Passive and active transport of ions through a membrane. Picture taken from [Ion].

Among irreversible processes in biology, *active* processes are those which use cellular energy in one or several instances along the process. A paradigmatic example of an active process is the flow of ions through membranes against the concentration gradient, which is mediated by ion pumps that use the energy from ATP hydrolysis. We notice that not all the biological irreversible processes are active. For example, diffusive transport through a membrane does not require any extra energy input and it can be seen as a passive nonequilibrium process in which an ion flow is produced by a concentration gradient.

Active processes are also responsible of biological processes involved in hearing [Hudo8]. Mechanical stimuli produced by sound are amplified in the inner ear. The amplification is an active process since the energy output is greater than the energy input, which is only possible if an amplifier contributes to the energy balance. The active "amplifier" are an ensemble of cells located in the cochlea, the ear hair cells, which use the ATP hydrolysis as energy input and whose structure and function we now discuss.

In vertebrates, the hearing process starts when the sound enters into the external auditory canal and hits the ear drum. When the ear drum is stimulated, the bones located in the middle ear (which is formed by the bones immersed in air) move in a synchronous way. The last bone hits another membrane connected to the cochlea, which is located in the inner ear and filled by an aqueous solution with a high concentration of potassium called *endolymph* (see Fig. 34 **a**). Inside the cochlea, the spiral organ of Corti is responsible of the transduction of the sound. Within the spiral organ of Corti, *hair cells* play a key role in the transduction of the sound since they transform the mechanical impulse into an electric signal that travels to the auditory nerves. Hair cells have a high frequency selectivity. Different hair cells are optimized to process sounds of different amplitudes and frequencies, and they are distributed in the cochlea according to their characteristic frequency as shown in Fig. 34 **b**. The structure of hair cells is shown in Fig. 34 **c**. An ensemble of actin filaments called *stereocilia* bend at the base of the apical surface of the hair cells and form the *hair bundles*. Hair bundles are stimulated by mechanical stimuli that enter into the cochlea and the collective oscillations of the stereocilia produce a flow of Ca^{2+} ions through the cell body which is the origin of the electrical impulse transmitted to the auditory nerves.

Ear Hair bundles are therefore the final responsible of the transduction of the sound and our object of study. They are formed by an ensemble of $20-300$ cylindrical actin stereocilia which are connected by flexible tip links, which are composed essentially by a protein called cadherin (see Figs. 35 **a-b**). Each stereocilium has ion channels that can be open or closed and are permeable to

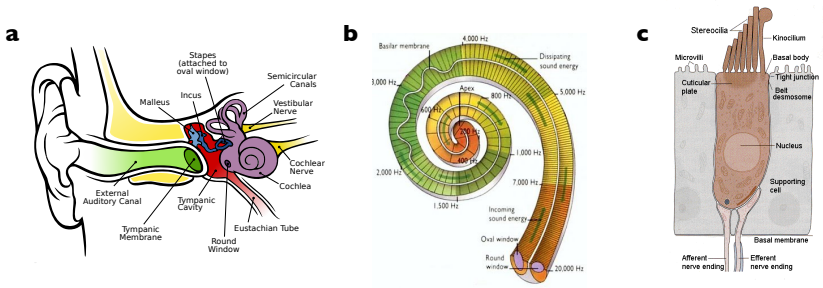


Figure 34: Auditory system in vertebrates. **a**. Human auditory system. Eardrum (green) hits the bones located in the middle ear (blue) which stimulate the oval window in the beginning of the cochlea (purple). Picture taken from [Ear]. **b**. Distribution of hair cells of different characteristic frequencies inside the cochlea. Picture taken from [Coc]. **c**. Ear hair cell. The ensemble of stereocilia above the basilar membrane form the hair bundle. Picture taken from [HbW].

Ca^{2+} ions. When the sound is propagated to the cochlea, stereocilia bend at their base and the hair bundle is deflected. This motion is accompanied by a shearing between stereocilia which is thought to open the ion channels. As a result, a current of Ca^{2+} ions towards the cell body is originated. Moreover, the Ca^{2+} concentration regulates the motion of myosin motors that can rectify or amplify the motion of each stereocilium in a mechanism called *adaptation* [Hudo8, Mano8].

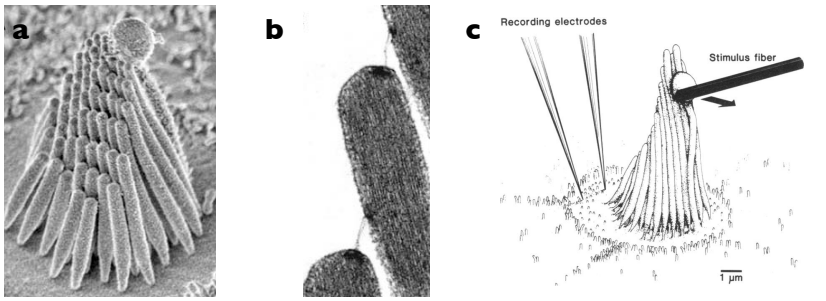


Figure 35: Microscopic structure of the ear hair bundle. **a.** Ear hair bundle of bullfrog's sacculus. The distance from bottom to the top is $\sim 6.7\mu$m. **b.** Magnified picture of two stereocilia connected by a tip-link. **c.** Scheme of a ear hair bundle controlled by a stimulus fiber where two electrodes are placed at its base to measure the transduction electric current. Pictures taken from [Mano8] and [How87].

From the point of view of biophysics, the dynamics of hair bundles has four main characteristics which can only be explained by the presence of active processes [Hudo8]:

- **Amplification** of mechanical stimuli has been observed in bullfrog's sacculus. Amplification of a sinusoidal stimulation of nanometer scale amplitude and frequencies of the

order of Hertz has been reported in bullfrog's sacculus hair bundles [Mar99]. The amplification factor can be even of order 10.

- **Frequency selectivity**: Different hair cells are adapted to different oscillation frequencies. Frequency tuning is very sharp in ear hair bundles, as it was shown in frogs [Maro1a], where hair cells tuned in the range $5-50$ Hz were observed in frogs.

- **Compressive nonlinearity**: The relation between sensitivity to an external force and the amplitude of the force is non linear. For example, in mammals, a relative increase of 1 million times in the amplitude of the stimulus increases the amplitude of the response of the hair bundle only by a factor 100 [Hudo8]. In frogs, the relation between sensitivity and amplitude of the stimulation is linear below amplitudes of 5nm and above 100nm but behaves like a power law of power $-2/3$ in the range $5-100$ nm [Maro1a].

- **Spontaneous otoacoustic emission**: Hair bundles can produce sound in a quiet environment as well as produce sounds at frequencies that are different to the stimulation frequency. Spontaneous oscillations of hair bundles has been observed in frog's sacculus [Maro3], and this motion has been discarded to be caused only by thermal fluctuations but also because of other active mechanisms [Maro1b].

These four characteristics of the dynamics suggest that the dynamics of the hair bundle can be described by a dynamical system operating near a *Hopf bifurcation*, which is a mathematical model that shares the aforementioned properties. A bifurcation occurs when a qualitative change on the behavior of a dynamical system changes when the value of one parameter of the system changes. Specifically, a *Hopf* bifurcation occurs when a stable point of a system transforms into a limit cycle with an unstable point inside the limit cycle [Stro1]. In other words, a system experiences a Hopf bifurcation when its dynamics changes from a damped oscillation (stable point) towards a sustained oscillation of constant amplitude (limit cycle) or vice versa. We notice that the Hopf bifurcation can be subcritical or supercritical, as illustrated in Fig. 36. Recently, both linear [Maro1b] and nonlinear [Nado4] models with a Hopf bifurcation have been used to describe the dynamics of hair bundles.

The biophysical description of the transduction of the sound at the molecular level can be done by taking into account mechanical properties of the main constituents of the hair bundle. Unlike the vision or the olfaction, sound amplification can be explained without the intervention of many intermediate biochem-

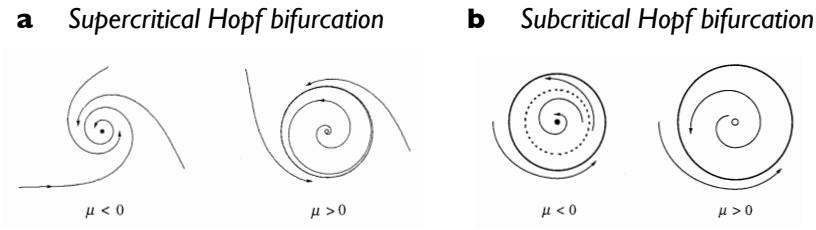


Figure 36: Supercritical and subcritical Hopf bifurcations. **a.** Example of a supercritical Hopf bifurcation described by the differential equations for the polar coordinates $\dot{r} = \mu r - r^3$; $\dot{\theta} = \omega + br^2$. The bifurcation occurs when μ changes from a negative value to a positive value. When μ changes from a negative value to a positive value, a stable fixed point turns into an unstable fixed point surrounded by a stable limit cycle. **b.** Example of a subcritical Hopf bifurcation described by $\dot{r} = \mu r + r^3 - r^5$; $\dot{\theta} = \omega + br^2$. When μ changes from a negative to a positive value a unstable limit cycle (dashed circle) shrinks into an unstable fixed point. The system oscillates when it reaches the stable outer limit cycle [Stro1].

ical reactions. Instead of that, the *gating-spring model* explains the transduction of the sound with the contributions of three mechanical elements: stereocilia, gating-springs and transduction channels [Mano8]. The basilar membrane of hair cells is resistant to stereociliar bending, which is quantified by a stereociliar pivot stiffness K_{sp}. gating-springs are mechanical elements that connect ion channels of different stereocilia. When the bundle deflects, gating-springs are stretched and transduction channels open. The resistance of gating-springs to stretch is measured by the gating-spring stiffness, K_{gs}. Tip links are thought to be a portion of each gating-spring. Gating of ion channels is stochastic and produces the shortening of the gating-spring by a few nanometers. This stochastic change of the force between stereocilia enters into the gating-spring model as a nonlinear term that is essential to explain the *gating compliance*, that is, the negative value of the gating-spring stiffness for small displacements of the bundle. Fig 37 sketches the gating-spring mechanism in hair bundles.

Apart from pure mechanical components, hair bundles need an active energy source in order to perform work. Motor proteins present inside the stereocilia are responsible of the active process that enhances the amplification of the sound [Hudo8]. Moreover, molecular motors play also a key role on the *adaptation* of the hair bundle, that is, the reset of the responsiveness of the hair bundle after strong stimuli. Actin filaments that form stereocilia can be stretched or compressed because of the action of myosin-1c adaptation motors as shown in Fig. 37**b**. The adaptation, and therefore motor activity, is strongly dependent on Ca^{2+} concentration, despite the molecular description of the Ca^{2+} is still not known

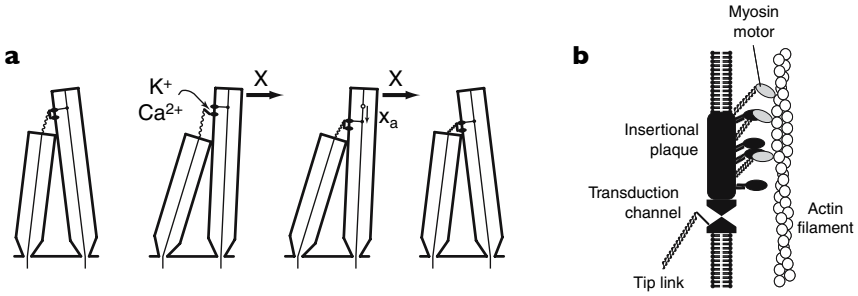


Figure 37: Gating-spring model of the hair bundle. **a.** Sketch of two stereocilia linked by a gating spring. When the two stereocilia are deflected, the gating-spring stretches and opens the transduction channel, which originates a current of ions towards the intracellular medium. The ion current acivates the adaptation motors that relax the spring to its original extension. **b.** Sketch of the structure and location of the adaptation Myosin motors that are responsible of the active processes in the hair bundles. Myosin motors are connected to the outer membrane of the stereocilia and move along actin filaments located inside stereocilia.

in detail. The simplest model [Mano8] expressed the active force exerted by the motors, f_a, in terms of the calcium concentration inside the hair cell C with a linear relation

$$f_a = f_{max}(1 - S\, P_o), \tag{5.1}$$

where P_o is the open channel probability, f_{max} is the maximum force that can be exerted by the motors, and S is the strength of Ca^{2+} in the motor force. S is defined as

$$S \equiv -\frac{C_M}{f_{max}} \frac{df_a}{dC}, \tag{5.2}$$

where C_M is the maximum Ca^{2+} concentration in the motor site.

The detection of active processes in hair bundles requires in principle knowing many degrees of freedom of the system, such as the position of the top of the hair bundle, the position of the collection of motor proteins or the Ca^{2+} concentration. Ear hair bundles can oscillate spontaneously in two different manners: first, they can oscillate due to the active processes such as motor activity, and secondly, they perform passive oscillations in the absence of any active process but due to thermal fluctuations of the environment. If one can only measure experimentally the position of the top of the bundle as a function of time, distinguishing between active and passive stochastic oscillations is not straightforward, since in both cases the trajectory of the top of the bundle is stochastic and oscillatory. However, in [Maro1b] Martin *et al.* were able to distinguish between active and passive oscillations in bullfrog's ear hair bundle using only the information of the

position of the top of the bundle. By taking into account the fluctuations of the position of top of the bundle in both spontaneous and forced oscillations, they measured the effective temperature that satisfies the fluctuation-dissipation relationship (FDR) (1.43). The temperature that the environment might have in order to sustain in equilibrium the correlations of the position of the top of the bundle (in Fourier space) $\tilde{C}_x(\omega)$ with a response to a weak external force $\tilde{\chi}(\omega)$ is

$$\frac{T_{eff}(\omega)}{T} = \frac{\omega \tilde{C}_x(\omega)}{2kTIm[\tilde{\chi}(\omega)]}. \tag{5.3}$$

This effective temperature is a measure of the degree of violation of the FDR. In [Maro1b] they found that the effective temperature of passive (dead) cells equals to the temperature of the environment, $\frac{T_{eff}(\omega)}{T} \simeq 1$ for all frequencies ω whereas the effective temperature of active (alive) cells does not $\frac{T_{eff}(\omega)}{T} \neq 1$. Moreover, in the active case, the dependence of the effective temperature on the frequency can be fitted to a linear stochastic oscillator model operating near a Hopf bifurcation.

In the next sections, we develop a method to distinguish between active and passive oscillations in ear hair bundles by using the information of the position of the top of the bundle recorded *only* from spontaneous − but not from forced − oscillations. We use the KLD rate of trajectories of the hair bundles from both experiments and simulations to do this distinction.

5.2 SIMULATIONS

We first check if we can distinguish between active and passive oscillations by measuring the KLD rate between forward and reverse simulated trajectories of the top of the hair bundle. We simulate the dynamics of the hair bundle by using the nonlinear stochastic oscillator model introduced by Nadrowski *et al.* in [Nado4]. Nadrowski's model has the main characteristics of the gating-spring model discussed in the previous section and reproduces with high accuracy the experimental values of relevant physical magnitudes of the hair cells, such as the characteristic frequency of the spontaneous oscillations or the sensitivity to an external force. The model is valid for hair cells with characteristic frequencies on the order of magnitude ranging from 1 to 100Hz. The model is described by three variables: the position of the top of the bundle, X, a second variable that plays the role of

the position of the collection of motors, X_a, and the intracellular Ca^{2+} concentration, C. The following set of equations

$$\lambda \frac{dX}{dt} = -K_{gs}(X - X_a - DP_o) - K_{sp}X + \eta, \tag{5.4}$$

$$\lambda_a \frac{dX_a}{dt} = K_{gs}(X - X_a - DP_o) - \gamma N_a f p(C) + \eta_a, \tag{5.5}$$

$$\tau \frac{dC}{dt} = -C + C_M P_o + \delta c. \tag{5.6}$$

reproduce the spontaneous oscillations of the hair bundle in the absence of any external force. The spontaneous oscillations are reproduced in the variable X for some specific values of the parameters of the model [Nado4] which we now define. All the ingredients of the gating-spring model discussed in Chapter 5.1 are included in this nonlinear stochastic oscillator model as we now discuss.

The dynamics of X is governed by Eq. (5.4), where λ is the friction coefficient associated to X. K_{gs} and K_{sp} are gating-spring and stereociliar pivots stiffnesses, respectively. $D = d/\gamma$ is the displacement of the bundle's top due to the opening of a channel, which causes an extension of the gating spring by an amount d, γ being a geometrical factor relating d and D. P_o is the probability of the transduction channels to be open, which is assumed to be nonlinear [Mano8] $P_o = (1 + Ae^{-(X-X_a)/\delta})^{-1}$, where A and δ are two constants defined in [Nado4]. The term $-K_{gs}(X - X_a - DP_o)$ accounts for the deflection of the top of the bundle due to the change of the gating-spring extension. Notice that $K_{gs}DP_o$ is a term that depends on the open channel probability and is positive, therefore it can model the gating compliance phenomenon. The term $-K_{sp}X$ accounts for the resistance of the stereociliar pivots to bend at their base. Tast term in (5.4) models the thermal noise, that is, $\eta(t)$ is a Gaussian white noise with zero mean and correlation $\langle \eta(t)\eta(t') \rangle = 2kT\lambda\delta(t - t')$.

Equation (5.5) describes the evolution of X_a. λ_a is the slope of the force-velocity curves of the motors, N_a is the total number of motors present in the hair bundle, f is the force exerted by a single motor and $p(C)$ is the probability for an adaptation motor to be bound to actin, which is assumed to be linear, $p(C) \simeq p_0 + p_1 C$. The first term in (5.5), $K_{gs}(X - X_a - DP_o)$, models the change of X_a produced by the change in the gating-spring extension. The second term in (5.5) models the regulation of the motor activity by Ca^{2+} and is analogous to Eq. (5.1) by identifying the maximum force of the motors with $f_{max} = \gamma N_a f$ and the strength of the Ca^{2+} feedback with $S = -C_M p_1/p_0$. The noise term for X_a does not satisfy fluctuation-dissipation theorem despite it is also Gaussian with zero mean and correlation. The noise intensity is in this case $\langle \eta(t)\eta(t') \rangle = 2kT_a\lambda_a\delta(t - t')$, with $T_a = 1.5T$. The

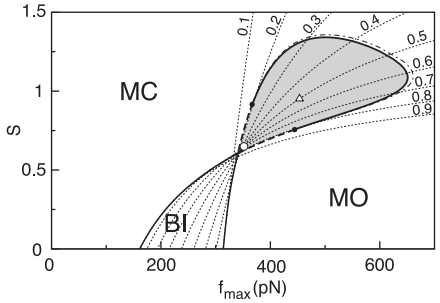


Figure 38: State diagram in the absence of noise in the Nadrowski model as a function of f_{max} and S. Spontaneous oscillations occur for parameter values in the dark shaded area. Dashed curves indicate the values of the parameters for which the open channel probability takes a certain constant value. The dashed boundary separating the oscillatory from the monostable (open and closed) regions corresponds to a subcritical Hopf bifurcation whereas the solid boundary represents the region where the Hopf bifurcation is supercritical. Picture taken from [Nado4]

fact that $T_a > T$ which accounts for an active mechanism that serves as energy source to perform work.

On the other hand, Eq. (5.6) describes the relaxation dynamics of the intracellular Ca^{2+} concentration, which is assumed to be initially zero. The relaxation time is denoted by τ, and C_M is the maximum concentration of calcium at the location of the motors. δc takes into account the fluctuations in the dynamics of Ca^{2+}, δc being a Gaussian white noise of zero mean and correlation $\langle \delta c(t)\delta c(t')\rangle = (1/2)C_M^2 N^{-1}\tau_c \delta(t-t')$, where N is the number of stereocilia and τ_c is the dwell time of the transduction channels. For the values of the parameters that fit the spontaneous oscillations of bullfrog's hair bundles, one can define an effective temperature in the intensity of calcium concentration fluctuations which is of order $T_c/T \simeq 0.13$, suggesting that Ca^{2+} fluctuations are one order of magnitude below thermal fluctuations [Nado4].

For the specific case of bullfrog's hair bundles, the majority of the parameters listed are known from experimental observations (see table I in [Nado4]). There are only two free parameters on the model: the maximal force produced by the motors in the direction of the movement, $f_{max} = N_a f p_0$, and the feedback strength of the calcium regulation, defined as $S = -C_M p_1/p_0$. In the absence of noise, the oscillator described by this model exhibits different dynamical behaviors depending on the values of f_{max} and S as shown in Fig. 38.

- **Monostable closed (MC)**: At low values of the motor force, adaptation motors cannot open the transduction channels and the majority of them are closed. This is also the case

when calcium feedback is very high and motors cannot exert forces strong enough to open the channels.

- **Monostable open (MO)**: When the force exerted by the motors is very large, the majority of the channels are maintained open.
- **Oscillatory (O)**: At intermediate values of Ca^{2+} feedback and motor forces indicated in Fig. 38 the system oscillates spontaneously.
- **Bistable (BI)**: Also at intermediate values of calcium feedback and motor forces, but in a different region in which calcium feedback is weaker than when oscillations occur, the system is in a bistable region where open and close channels coexist.

In the case of bullfrog's ear hair bundle, the *operating point*, that is, the values of the parameters that reproduce Bullfrog's bundle's spontaneous oscillations, is located at $f_{max} = 325$pN and $S = 0.65$ (circle mark in Fig. 38), which is very close to the boundary between the oscillatory region and the bistable region. When crossing the boundary between the different regions, the system experiences a Hopf bifurcation. Notice that the bifurcation is subcritical near the operating point (dashed boundary in Fig. 38) and it becomes supercritical for large values of f_{max} (solid boundary in Fig. 38). The values of the parameters for which the sensitivity of the system to an external force is maximum is also indicated in Fig. 38 with a triangular mark.

We simulate the model described by Eqs. (5.4)-(5.6) for different values of f_{max} and S using Euler's numerical integration scheme described in Appendix D.1. We first analyze the dynamics of the system when it crosses the bifurcation in the region where it is supercritical. We vary the parameters following the values indicated in the left panel in Fig. 39. For each point (f_{max}, S) in the curve indicated in the figure, we simulate a single trajectory of $n_0 = 10^8$ data points with time step $\Delta t_0 = 0.012$ms, that is, simulations of total time $T_0 = n_0 \Delta t_0 = 1200$s. From the simulations, we only sample the position of the top of the bundle, $(X, X_a, C) \to X$, using a sampling frequency of $f_s = 1/(10\Delta t_0) = 8.3$kHz (*i.e.* We sample the trajectories $X(t)$ every 10 data). In the end, we work with time series of $n = 10^7$ data points with time step $\Delta t = 0.12$ms and total time $T = n\Delta t = 1200$s. For every trajectory, we measure the KLD rate *per unit of time* with the estimation technique described in sec. 3.2.2 [Eq. (3.32)]. In the estimation, we first fit the data to an $AR(10, 50)$ model, that is, an $AR(k, l)$ model of order $k = 10$ and lag $l = 50$, and calculate the KLD using the residual associated to this AR model. This procedure is equivalent to consider blocks of 10 data separated

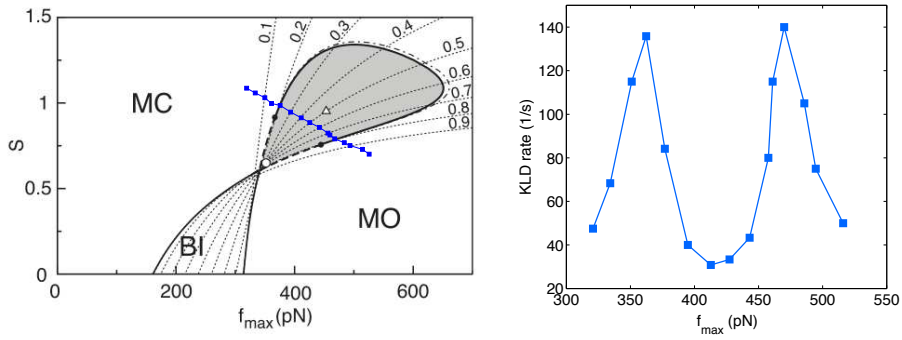


Figure 39: KLD rate of trajectories of the top of the bundle when crossing a supercritical Hopf bifurcation in Nadrowski's model. Left: Variation of the simulation parameters along a supercritical Hopf bifurcation. Right: KLD rate of single stationary trajectories of the position of the top of the hair bundle, $X(t)$, with total time of the trajectory $T = 1200s$ and $f_{acq} = 8.3kHz$, $\hat{d}^x_{AR}$, as a function of f_{max} along the line depicted in the left panel.

$t_{lag} = l\Delta t = 6ms$, $t_{block} = kt_{lag} = 60ms$ being the total time of the block, which is much smaller than the period of the oscillations $t_{block} \ll t_{osc} = O(1s)$.

We show in the right panel of Fig. 39 the value of $\hat{d}^x_{AR}$ (3.32) as a function of f_{max} along a curve that crosses the supercritical bifurcation shown in the right panel of Fig. 39. Our estimator increases when approaching the oscillatory region and has two maxima that occur when the system crosses the supercritical Hopf bifurcation. This result suggests that the time irreversibility in X is maximum in the boundary between the oscillatory and the monostable (open or closed) regions, and therefore KLD can be used to detect the presence of a bifurcation.

An analogous result is obtained when the bifurcation is crossed with a similar curve but in the region where the Hopf bifurcation is subcritical. We simulate trajectories of the same time length and sampling frequency than in the supercritical case. Similarly to what happens in the supercritical Hopf bifurcation, the KLD rate calculated using the position of the top of the bundle $\hat{d}^x_{AR}$ increases when the system approaches the oscillatory region from any monostable region and decreases within the oscillatory region the farther the system is from the bifurcation (see Fig. 40). If the bifurcation is crossed twice two maxima are observed in the vicinity of the subcritical bifurcation as it happens in the supercritical bifurcation [cf. Fig. 39].

We now analyze the behavior of the KLD when we vary the parameters following a path with constant open channel probability close to $P_o \simeq 0.5$ in the state diagram in the absence of noise (see left panel of Fig. 41). For every (f_{max}, S) point in the curve of the left panel of Fig. 41 we simulate the hair bundle

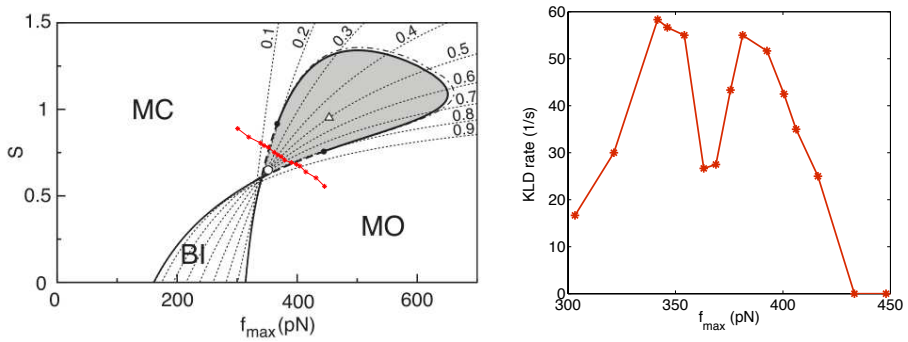


Figure 40: KLD rate of trajectories of the top of the bundle when crossing a subcritical Hopf bifurcation in Nadrowski's model. Left: Variation of the simulation parameters along a subcritical Hopf bifurcation. Right: KLD rate of single stationary trajectories of the position of the top of the hair bundle, $X(t)$, with total time of the trajectory $T = 1200s$ and $f_{acq} = 8.3kHz$, $\hat{d}^x_{AR}$, as a function of f_{max} along the line depicted in the left panel.

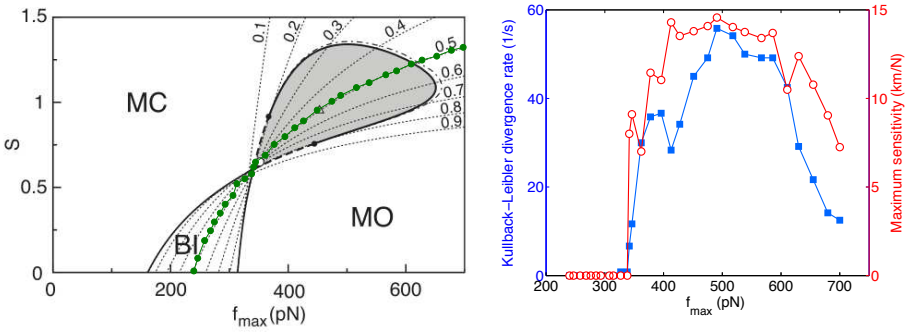


Figure 41: KLD rate of trajectories of the top of the bundle in the bistable to oscillatory transition in Nadrowski's model. Left: Variation of the simulation parameters along a curve with constant open channel probability $P_o = 0.5$. Right: KLD rate per unit of time $\hat{d}^x_{AR}$ (blue squares, left axis) and maximum sensitivity (red circles, left axis) as a function of f_{max} for single stationary trajectories of $T = 1200s$ and $f_{acq} = 8.3kHz$ with the parameters given by the values in the left panel.

dynamics using Nadrowski's model and generate trajectories of 1200s with simulation time step of $\Delta t_0 = 0.012ms$. We restrict ourselves to the information given by the position X, and sample the trajectories at $f_s = 8.3kHz$, obtaining, for each value of the parameters, a single stationary trajectory $X(t)$ that contains $n = 10^7$ data points each of them sampled every $\Delta t = 0.12ms$. We measure the KLD rate estimator $\hat{d}^x_{AR}$ as well as the maximum sensitivity of the hair bundle for every point in the parameter space curve. When increasing the value of S, the system crosses the bifurcation going from the bistable to the oscillatory regime. For the values of f_{max} and S indicated in the left panel of Fig. 41, $\hat{d}^x_{AR}$ increases with the force for moderate values of f_{max} and it exhibits a maximum that occurs near the maximum sensitivity point.

We now focus on the problem of distinguishing between active and passive oscillations of the hair bundle using only the information from the position of top of the bundle. In this case, $X(t)$ is also a partial description of the dynamics, and it is not straightforward to guess from the time traces if they are generated by an active or a passive oscillator. We first notice that, if we neglect calcium dynamics, Nadrowski's model can be expressed in terms of a potential $V(X, X_a)$ that depends on the variables X and X_a,

$$\lambda \frac{dX}{dt} = -\frac{\partial V(X, X_a)}{\partial X} + \eta, \tag{5.7}$$

$$\lambda_a \frac{dX_a}{dt} = -\frac{\partial V(X, X_a)}{\partial X_a} - \gamma N_a f p(C) + \eta_a. \tag{5.8}$$

Therefore the majority of the terms in Nadrowski's model are conservative. The non conservative term, $-\gamma N_a f p(C)$, accounts for the active process generating the spontaneous oscillations. As a consequence, one can simulate passive oscillations by setting $f_{max} = 0$, which makes the active term to vanish. However, the richness of this model allows us to analyze the system in a variety of situations aside from the active and passive cases. We now study how the KLD rate per unit of time scales with time for the different regions of the state diagram of Nadrowski's model and whether or not we can distinguish between them by means of the behavior of the KLD. In Fig. 42 we sample the position of the top of the bundle $X(t)$ from trajectories simulated using Nadrowski's model. In all cases, we generate trajectories with $\Delta t_0 = 0.012ms$ and sample the trajectories with an acquisition frequency of $f_{acq} = 8.3$ kHz. In Fig. 42 we show simulated trajectories corresponding to three different situations. In the top and the middle panel of the figure we show the position of the top of the bundle obtained from trajectories generated in the monostable region where the majority of the ion channels are open or closed (MO and MC). In these regions, the oscillator does not perform spontaneous oscillations similar to the ones ob-

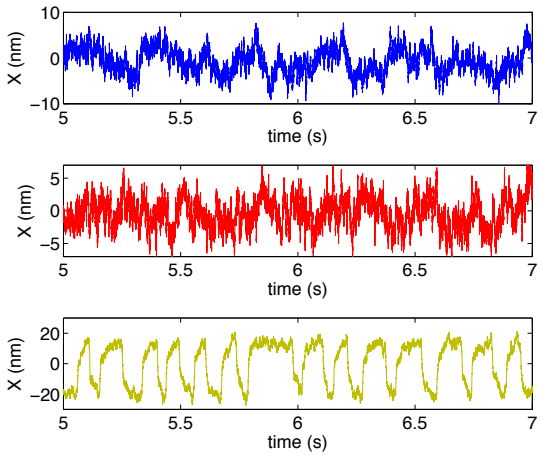


Figure 42: Position of the top of the hair bundle (in nm) as a function of time (in seconds) in simulations of the Nadrowski's model. Top: $f_{max} = 100$pN and $S = 0.5$, corresponding to a monostable closed situation. Medium: $f_{max} = 300$pN and $S = 0.25$ corresponding to a bistable oscillation. Bottom: $f_{max} = 325$pN and $S = 0.65$, corresponding to the operating point of the hair bundle.

served experimentally [Maro3]. The third situation (bottom panel in Fig. 42) corresponds to the operating point of the hair bundle, where spontaneous oscillations are produced by active processes. Statistically speaking, the latter time series is in appearance more irreversible than the other two because of the presence of irreversible patterns in the series. More specifically, in the oscillation shown in the bottom panel of Fig. 42, slow excursions in the negative direction of X are interrupted by fast jumps in the positive direction in accordance to the structure of the spontaneous oscillations observed in bullfrog's hair bundles [Maro3].

We study how our KLD rate estimator $\hat{d}^x_{AR}$ scales with time when we consider time series of different total time t, $\{X(t')\}^t_{t'=0}$, in an equivalent analysis to the one performed in chapter 3.3, where we used the scaling KLD estimator based on visibility algorithm to distinguish between equilibrium and NESS. We denote by $\hat{d}^x_{AR}(t)$ the KLD rate estimator using the residual of the $AR(k,l)$ model calculated using a single stationary trajectory of the top of the bundle of total time t, . We investigate the behavior of $\hat{d}^x_{AR}(t)$ for different regimes of the oscillator by using the same estimator as before, measuring the residual of an $AR(10,50)$ model. For that, we generate trajectories of $T = 125$s, with $\Delta t_0 = 0.012$ms and sampled with $f_{acq} = 8.3$ kHz, and measure $\hat{d}^x_{AR}$ of subsequences of total time $t < T$, starting at $t' = 0$ and ending in t, $\{X(t')\}^t_{t'=0}$. In Fig. 43 we show the value of $\hat{d}^x_{AR}(t)$ as a function of time t in double logarithmic scale for different

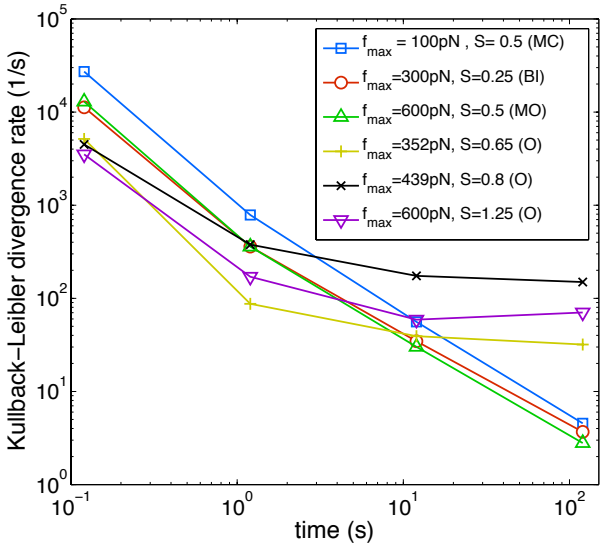


Figure 43: Scaling of the KLD rate with time of the trajectory in the different regions of Nadrowski's model. KLD rate as a function of time $\hat{d}^{x}_{AR}(t)$ using the residual of an $AR(10,50)$ model for trajectories of the bundle's top $X(t)$ as a function of the total time of the trajectory (in s). For each of the 6 cases shown in the figure, we simulate a single stationary trajectory of total time $T = 125s$ which is sampled at a rate of $\Delta t = 0.12ms$.

values of time $t = 0.125s$, $1.25s$, $12.5s$ and $125s$. The scaling of $\hat{d}^{x}_{AR}(t)$ with time is different depending on which region of the state diagram the oscillator is, as shown in Fig. 43. If the oscillator is in the monostable state (both open or closed) or in the bistable state, $\hat{d}^{x}_{AR}(t)$ scales like $1/t$ and tends to 0 when t is large (see Fig. 43). Conversely, in the oscillatory regime, $\hat{d}^{x}_{AR}(t)$ also decreases with time but it tends asymptotically to a constant an positive value when time increases. Therefore, only in the regime where spontaneous oscillations are reproduced, the KLD tends to a positive value which suggests that the system is in a NESS. Moreover, the value of the plateau of the KLD rate for long times is of order $100\ kT/s$. This result can be related to the average dissipation per unit of time of the system provided the bound (2.33). In this case, $\frac{\langle \dot{W}_{diss} \rangle}{kT} \geqslant \hat{d}^{x}_{AR}$, and $\hat{d}^{x}_{AR}$ tends to a value whose order of magnitude is $100\ s^{-1}$ which implies that the average dissipation of the oscillator is at least $100kT/s$. We notice that the ATP consumption rate for a myosin adaptation motor is of order $r_{ATP} = 6s^{-1}$ (see Table 13.1 in [Howo1]) and therefore the energy consumption rate of the ATP hydrolysis is of order

$$\dot{W}_{diss} \simeq \Delta G_{ATP} \times r_{ATP} = \frac{25kT}{ATP} \times \frac{6ATP}{s} = 150\ kT/s. \qquad (5.9)$$

Consequently, our KLD rate estimator predicts that the minimum irreversibility to sustain spontaneous oscillations in Nadrowski's

model is of the order of magnitude of the dissipation rate of a single adaptation motor.

We notice that the KLD rate measured using the information of the position can be misleading when distinguishing between active and passive processes. Passive oscillations are reproduced in Nadrowski's model when $f_{max} = 0$, which is in the MC region, where the KLD vanishes. However, not all the active oscillations $f_{max} > 0$ yield a positive value of the KLD rate for trajectories of the bundle's top, and in some cases such as the blue, red and green curves in Fig. 43 one might think that oscillations are passive since $\hat{d}^{x}_{AR}(t) \to 0$ when t is large. Therefore, only $\hat{d}^{x}_{AR}(t)$ tends to a positive value we can ensure that the trajectories generated by the Nadrowski's model are generated by underlying active processes. This case is similar to a discrete flashing ratchet with infinite switch rate, where, even using full information of the system, the KLD rate vanishes whereas the dissipation does not [Rol10].

In the next section, we extend our analysis to experimental data obtained from bullfrog's ear hair bundles. We try to ascertain if we can detect active processes by measuring the KLD rate of trajectories of the top of the hair bundle.

5.3 APPLICATION TO EXPERIMENTAL DATA

We now study if our technique can be applied to distinguish between active and passive oscillations of hair bundles using experimental data from bullfrog's sacculus. The experimental data was obtained by J. Barral and P. Martin in *Laboratoire de Physico-Chimie Curie (Institut Curie)*[2]. The data was obtained with the same protocol as that described in [Mar03]. The ear is dissected from an American bullfrog (*Rana catesbeiana*) after the animal has been sacrificed (a needle is inserted into the brain from the back of the head). Bullfrog's sacculus is mounted in a two-compartment chamber where cell bodies are immersed in a saline solution and hair cells in a second compartment filled by an artificial endolymph. Experiments are done at a room temperature $T \sim 21°C$. Video microscopy is used to find active hair bundles that oscillate spontaneously. A flexible glass fiber of $\sim$ 500 nm diameter is attached to the top of the bundle as sketched in Fig. 35 **c** and monitored at a magnification of $1000\times$ on a dual photodiode. The displacement monitor provides a linear relationship between the displacement of the tip of the fiber and the voltage of the photodiode which allows to measure the position of the top of the bundle with an accuracy of $\sim$ 1 nm.

2 Full address: Laboratoire de Physico-Chimie Curie, Unité Mixte de Recherche 168, Institut Curie, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

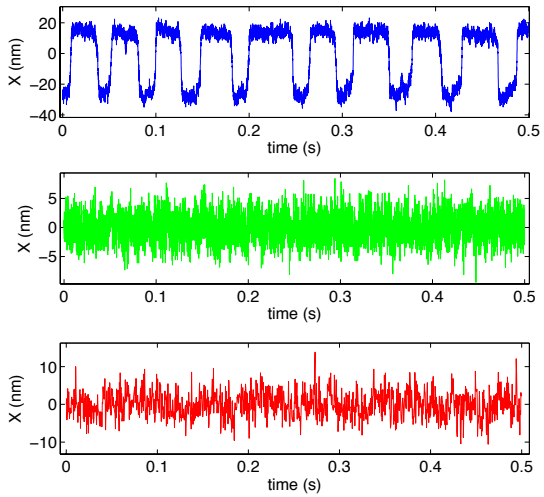


Figure 44: Displacement of the top of the bundle with respect to its average value (in nm) as a function of time (in s) for different cells. Top: Hair cell that exhibits active oscillations, recorded with an acquisition frequency of 14kHz during 90s of experiment. Middle: Same cell after the exposure to gentamicin also sampled at 14kHz. Bottom: Passive cell displacements, sampled at 2.5kHz for 60s of experiment.

Using the experimental procedure described above, we monitored the position of the top of the bundle in spontaneous oscillations that lasted $\sim$ 100 s with a peak-to-peak amplitude of $\sim$ 10nm. We use the data of different hair cells that oscillate with different characteristic frequencies of the order of Hz. Sampling rate was of the order of kHz. On the other hand, oscillations of dead cells and experimental noise were also recorded to check if we can distinguish them from the oscillations of active (living) cells by means of the KLD rate of the trajectories of the position of the top of the bundle.

First, we measure the position of the top of a hair bundle $X(t)$ of a cell that exhibits spontaneous oscillations of amplitude $\sim$ 20nm using an acquisition frequency of 14kHz (see top panel in Fig. 44). Secondly, we added gentamicin to the endolymph of the chamber where hair cells are exposed. As indicated in [Mar03], gentamicin blocks transduction channels and produces that the oscillations of the bundle become Hookean of amplitude $\sim$ 5nm as shown in the middle panel of Fig. 44. As a third case study, we use the data of hair cells recorded several minutes after performing spontaneous oscillations. After this time, hair cells die and oscillate in a passive manner because of thermal fluctuations. In this case, the position of the bundle is also stochastic and oscillatory with an amplitude of $\sim$ 10nm as shown in the bottom panel of Fig. 44.

We now analyze how the KLD rate of a single stationary trajectory of the top of the bundle $X(t)$ — obtained in a single

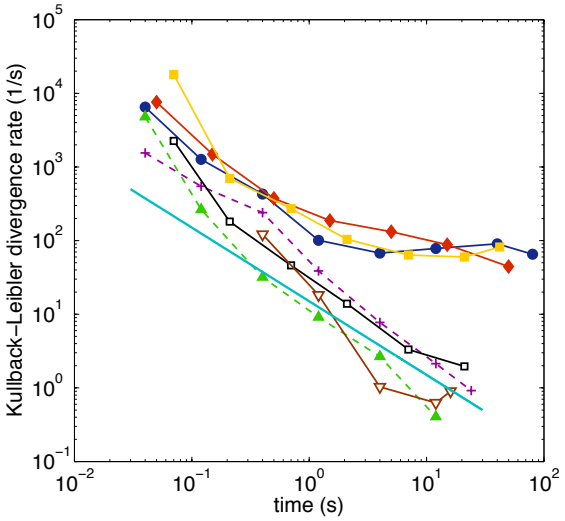


Figure 45: Scaling of the KLD rate estimator $\hat{d}^{x}_{AR}$ with time for spontaneous oscillations of the top of the bundle obtained from bullfrog's sacculus. We measure the estimator of the KLD rate defined in Eq. (3.32) for subsequences of time t of a single experiment, $X(t')_{t'=0}^{t}$, $\hat{d}^{x}_{AR}(t)$, using the residual of an $AR(10,50)$ model. Yellow squares: Cell 1, with total time of the recording $T = 60$s and data acquisition rate $f_{acq} = 14$kHz. Blue circles: Cell 2, $T = 90$s and $f_{acq} = 25$kHz. Red diamonds: Cell 3, $T = 50$s and $f_{acq} = 20$kHz. Black open squares: Cell 1 with same total recording time and acquisition frequency in gentamicin endolymph. Brown: Passive cell, with $T = 30$s and $f_{acq} = 2.5$kHz. The KLD of the experimental noise is also shown in dashed lines (Magenta "+" noise light with $T = 30$s and $f_{acq} = 25$kHz, green triangles noise off with $T = 30$s and $f_{acq} = 25$kHz). In cyan, we add a line that has a dependence of $1/\text{time}$ to guide the eye.

experiment – scales with time for different hair cells. We use the same estimator of the KLD as that used in simulations in chapter 5.2, that is, $\hat{d}^{x}_{AR}$ using the residual of an $AR(10,50)$ model. Apart from the cells whose trajectories are sampled in Fig. 44, we also consider cases where we measured the position of the bundle in two other cells that oscillate spontaneously at different frequencies. For completion, we also recorded the experimental noisy signal from the photodiode (noise off? on?). Fig. 45 is a summary of our results where we show the value of $\hat{d}^{x}_{AR}$ for the different cases as a function of time t (details of every cell are specified in the caption). At every time t indicated in the figure, the KLD rate estimator is measured for a subsequence of a single experiment of total time t_{exp} starting in $t' = 0$ and ending in $t' = t < t_{exp}$. We observe that for the three cases in which the cell oscillates spontaneously, the KLD rate saturates for long times to a positive value that is greater than zero. The KLD of the rest of the signals, including the passive oscillation,

experimental noises and the oscillation of one of the cells in the presence of gentamicin, tend to zero like $1/t$ as shown in Fig. 45. We can discriminate between oscillations generated with active and passive processes using the scaling of $\hat{d}^x_{AR}$ with time since it tends to zero for passive processes and to a positive value for oscillations produced by active processes.

In the active case, the value at which the KLD rate saturates for active cells is of order $\hat{d}^x_{AR} \sim 100 s^{-1}$, which corresponds to an average dissipation per unit of time of $\langle \dot{W}_{diss} \rangle \sim 100 kT/s$ as discussed in the previous section. This value bounds from below the average energy consumed from ATP hydrolyzation per unit of time of a single adaptation motor as proved before in Eq. (5.9). Therefore using the KLD rate of trajectories of the top of the bundle we can predict the minimum irreversibility needed to sustain spontaneous oscillations in bullfrog's ear hair bundle.

5.4 CONCLUSIONS

In this chapter, we have applied our estimator of the KLD rate that uses auto regressive (AR) models to measure the time irreversibility of real-valued time series. Biological systems in the stationary state produce stochastic time series whose time irreversibility can be measured with our estimator of the KLD rate and relate its value with the dissipation rate or the ATP consumption rate of the system. In particular, we have applied our technique to distinguish between oscillations produced by active and passive processes in ear hair bundles.

We have first applied our technique to data produced by a nonlinear stochastic model of the hair bundle. We have measured the KLD rate of stationary trajectories of the top of the simulated hair bundle. Although the position of the top of the bundle is a partial description of the system, our estimator of the KLD rate detects when the system crosses a Hopf bifurcation when varying the parameters of the model. Moreover, we are also able to distinguish between different dynamical behaviors, such as oscillatory or monostable, in terms of the scaling our KLD rate estimator. In simulations, our measure of irreversibility underestimates the dissipation of active oscillations in the monostable region, whereas it correctly predicts the average ATP consumption rate of active oscillations.

We have also applied the same technique to experimental data of hair bundles extracted from bullfrog's sacculus. We have measured the position of the top of hair bundles in spontaneous oscillations in both active (alive) and passive (dead) hair cells. With our KLD rate estimator we are able to discriminate between active and passive hair cells of different frequencies and sampled with different data acquisition frequencies. Moreover, we are able

to predict the minimum irreversibility to sustain bullfrog's hair bundle spontaneous oscillations, which is very close to the energy consumption rate by a single myosin adaptation motor, which is the system responsible of the underlying active process.

Part III

EXPERIMENTAL TESTS AND APPLICATIONS OF STOCHASTIC THERMODYNAMICS

6

ENERGETICS OF THE SYMMETRY BREAKING

A symmetry breaking (SB) involves an abrupt change in the set of microstates that a system can explore. This change has unavoidable thermodynamic implications. According to Boltzmann's microscopic interpretation of entropy, a shrinkage of the set of compatible states implies a decrease of entropy, which eventually needs to be compensated by a dissipation of heat and consequently requires work. Examples are the compression of a gas and the erasure of information. On the other hand, in a spontaneous SB the available phase space volume changes without the need of work, yielding an apparent decrease of entropy. Here we show that this decrease of entropy is a key ingredient in the Szilard engine and the Landauer's principle and report a direct measurement of the entropy change along SB transitions in a Brownian particle. The SB is induced by a bistable potential created with two optical traps. The experiment confirms the theoretical results based on fluctuation theorems, allows us to reproduce the Szilard engine extracting energy from a single thermal bath, and shows that the signature of a SB in the energetics is mensurable, providing new methods to detect the coexistence of metastable states in macromolecules.

Landauer's principle is one of the first theoretical results that impose a limitation on small computing machines, proving that the erasure of a bit requires at least an energy of amount $kT \ln 2$ [cf. Eq. (1.16)]. Landauer principle was tested experimentally recently [Bér12] and it was shown that it can be obtained as a particular case of the KPB theorem relating irreversibility and dissipation [Kaw07]. The energetics of the erasure of a bit can be seen as the restoration (SR) of a broken symmetry where a Brownian particle has chosen a particular option. Because of this, the energetics of the erasure of a bit and/or any generic symmetry breaking (or restoration) in the microscopic scale can be studied in the context of our work. In this chapter we focus on the energetics of the SB and SR, finding a formula that relates the average entropy production when a microscopic system chooses a particular instance in a SB (or when the symmetry is restored from that option) with the probability to choose that option. Our formulas are at the core of the thermodynamics of choice and they are universal, *i.e.*, they do not depend on the mechanism that biases any of the two options. Interestingly, both Landauer principle and the energetics of the Szilard engine can be reproduced with

our formulas, and also the recent results on the generalization of the Second Law to feedback processes [Sag08, Sag09].

We design an experiment to test our formulas, where a Brownian microscopic dielectric particle is trapped with a dual optical tweezer that creates a double well potential. We produce a SB by separating the traps by moving one of them at constant velocity. By applying an external uniform electric field, we can tune the probability of the particle to choose between the two traps. Additionally, a SR is implemented with the reverse process, by rejoining the traps. As we discuss in this chapter, our experimental results not only validate our formulas but also serve to design a new experimental protocol to implement a Szilard engine simpler to the first microscopic design [Toy10].

This chapter is organized as follows: Chapter 6.1 is an introduction where we discuss the thermodynamic consequences of a symmetry breaking and the main theoretical and experimental previous results in literature that discuss this issue. In Chapter 6.2, we explore the connection between the main formula of the work (relating irreversibility and dissipation in the microscopic scale) with the energetics of a symmetry breaking and a symmetry restoration. We obtain universal expressions that express the energetics of a SB (or SR) with the probability of a microscopic system to choose a particular instance i. In Chapter 6.5 we report the experimental validation of our theoretical results obtained using optical tweezers. In Chapter 6.6 we show how our setup can be used to design a Szilard engine. The discussion of the results is done in Chapter 6.7.

6.1 IRREVERSIBILITY AND SYMMETRY BREAKING

When a symmetry is broken, a system "makes a choice" among a set of instances $i = 1, \dots, m$. For a classical infinite system, a symmetry breaking (SB) consists of a sudden change in the set of available states: the whole phase space Γ is partitioned into non overlapping regions Γ_i, corresponding to the different instances $i = 1, \dots, m$. The partition occurs when certain control parameter λ crosses a critical value λ_c above which the system can no longer move spontaneously from one region to another and gets confined within Γ_i with a probability p_i, $\sum_i p_i = 1$. The notion of SB can be extended to finite systems with metastable states. The confinement is not strict in this case: the system can jump form a region Γ_i to another Γ_j. However, if the average residence time in each region is much larger that the time scale of the process under consideration one can talk about an effective SB. In this case, the SB transition is not localized in a single value of the control parameter λ but it is rather a continuous transition where metastable states develop.

The energetics associated to SB transitions and, in general, to the manipulation of metastable states has special relevance in a number of interesting physical situations, some of them realized experimentally in the last years. The original Szilard engine, a celebrated and refined version of the original Maxwell demon, can extract work from a single thermal bath using the information created in a SB [Lef92, Paro1, Toy10]. Landauer's principle accounts for the minimum dissipation associated to the erasure of information, which is a manipulation of the two metastable states making up a single bit memory [Lef92]. The erasure can be interpreted as the restoration of a broken symmetry (see below) and has been reproduced with a Brownian particle in a double well potential created by optical tweezers [Bér12]. In molecules, metastable states correspond to different molecular conformations as well as to kinetic states of special relevance in biophysics. The energetics of processes involving metastable states has become a tool to measure conformational free energies in those contexts. An extended version of the nonequilibrium work theorem relates the probability distribution of the work in a process connecting two metastable states with their conformational free energies [Maro8]. Maragakis et al [Maro8] applied this result to numerical simulations of switches between two different conformations of alanine dipeptide and Alemany et al [Ale12] have used it to obtain conformational free energies in stretching experiments of DNA structures that posses intermediate and/or misfolded kinetic states.

In this chapter we report on an experimental realization of a SB consisting of a continuous transition from a single well to a double well potential affecting a Brownian particle. We reproduce the transition moving apart two optical traps and measure the heat dissipated by the particle to the surrounding water that acts as a thermal reservoir. An electrostatic field allows us to tune the bias towards one or the other trap and explore the relationship between the energetics of the SB and the probability to adopt one of the instances. We finally build a Szilard engine as a SB followed by the restoring of the symmetry. This process completes a cycle that extracts energy from the thermal bath if the electrostatic field along the process is properly chosen.

6.2 SYMMETRY BREAKING AND SYMMETRY RESTORATION

Consider a system with Hamiltonian $H(x;\lambda)$ ($x \in \Gamma$), depending on a control parameter λ, and an isothermal process at temperature T involving a SB, where the parameter changes in time as

λ_t with $t \in [0, \tau]$. The average work required to complete the process, when the system adopts instance i, is bound by

$$\langle W \rangle_i^{(SB)} - \Delta F_i \geqslant kT \ln p_i \tag{6.1}$$

where k is the Boltzmann constant and $\Delta F_i = F_{\tau,i} - F_0$ is the change in free energy. The initial free energy is defined as usual, $F_0 = -kT \ln Z(T, \lambda_0)$ where $Z(T, \lambda) = \int_\Gamma dx\, e^{-\beta H(x;\lambda)}$ is the partition function of the system. On the other hand, the final free energy $F_{\tau,i} = -kT \ln Z_i(T, \lambda_\tau)$ is a conformational free energy defined in terms of the partition function restricted to the region Γ_i, i.e., $Z_i(T, \lambda_\tau) = \int_{\Gamma_i} dx\, e^{-\beta H(x;\lambda_\tau)}$. The bound in equation (6.1) is met with equality if the process is quasi static. Recalling the relationship between the free energy, F, the internal energy E, and the entropy S of a system, $F = E - TS$, and the first law of thermodynamics $\Delta E = W + Q$, where Q is the heat or energy transfer from the thermal reservoir to the system, we easily derive a bound for the entropy production:

$$\langle S_{prod} \rangle_i^{(SB)} \equiv \Delta S_i - \frac{\langle Q \rangle_i^{(SB)}}{T} \geqslant kT \ln p_i \tag{6.2}$$

A rigorous proof of these bounds follows from fluctuation theorems (see Supplemental Material). However, the origin of the term $k \ln p_i$ in Eqs. (6.1) and (6.2) can be easily understood. A SB comprises a contraction of the set of available states from Γ to Γ_i without the need of any extra work. This amounts to an increase of free energy $-kT \ln(Z_i/Z)$ which is not compensated by work and heat dissipation. Assuming an instantaneous SB, $p_i = Z_i/Z$ yielding the extra term $kT \ln p_i$ in Eqs. (6.1) and (6.2).

This work-free shrinkage of the available phase space is entirely due to the SB transition and is not in contradiction with the Second Law of Thermodynamics, because the final state $\rho_i(x)$ is not in complete equilibrium and the final entropy cannot be considered as a true thermodynamic entropy. The conformational entropies S_i or free energies F_i, however, even though cannot be considered as true thermodynamic potentials (they are not state functions, for instance) are useful tools to analyze the energetics of processes involving SB transitions. An alternative interpretation of the decrease of entropy is that it is compensated by an increase of the meso- or macroscopic uncertainty, quantified by the Shannon entropy of the SB outcome, $H = -\sum_i p_i \ln p_i$. Notice that the average of Eq. (6.1) over p_i yields precisely $kT\,H$.

Similar inequalities hold for a process where a symmetry is restored. To assess the energetics of a symmetry restoration we have to consider the time reversal of the restoration, which is a SB. Let us call $\tilde{p}_i$ the probability that the system adopts instance i in this SB resulting from the time reversal of the original process. Under time reversal, reversible work and free energy increment

change sign. Therefore (see Supplemental Material for a rigurous proof):

$$\langle W\rangle_i^{(SR)} - \Delta F_i \geqslant -kT \ln \tilde{p}_i, \tag{6.3}$$

where $\Delta F_i = F_\tau - F_{0,i}$ is the free energy change of the symmetry restoration. Notice that now it is the initial free energy that depends on the instance i. For the entropy:

$$\langle S_{prod}\rangle_i^{(SR)} \geqslant -k \ln \tilde{p}_i. \tag{6.4}$$

Landauer's principle follows immediately from Eq. (6.3) applied to a one-bit memory consisting of a physical system with two stable states, 0 and 1, each one with the same free energy $F_0 = F_1$. The minimal cost of erasing a bit or, more precisely, to drive bit $i = 0$ or 1 to the state 0 (restore-to-zero operation) is $\langle W\rangle_i^{erasure} \geqslant -kT \ln \tilde{p}_i - \Delta F_i = -kT \ln \tilde{p}_i$ for $i = 0$ or 1, since in both cases $\Delta F_i = F_i - F_0 = 0$. If the initial bit is unknown, the best we can do is $\tilde{p}_i = 1/2$ and $\langle W\rangle_i^{erasure} \geqslant kT \ln 2$.

The energetics of the Szilard engine [Lef92, Sag09] can be as well easily reproduced from Eqs. (6.1) and (6.3). In the Szilard setup a system undergoes a SB and chooses between two instances 0 or 1 with probability p_0 and p_1, respectively. Then we measure the instance that has been chosen and restore the broken symmetry driving the system back to the original state through some protocol λ_t^i. The time reversal of this protocol is a SB transition with possibly different probabilities $\tilde{p}_j^i$. The work necessary to implement the SB is bound by Eq. (6.1) and the work necessary to restore the symmetry is bound by Eq. (6.3). Therefore, the total average work that we have to perform to run the whole cycle obeys:

$$\begin{aligned} \langle W\rangle &= \sum_i p_i \left[\langle W\rangle_i^{(SB)} + \langle W\rangle_i^{(SR)}\right] \\ &\geqslant kT \sum_i p_i \ln \frac{p_i}{\tilde{p}_i^i} \end{aligned} \tag{6.5}$$

and $\langle S_{prod}\rangle = \langle W\rangle/T$. In the case of the original Szilard engine: $p_i = 1/2$ and $p_i^i = 1$, yielding $\langle W\rangle \geqslant -kT \ln 2$, i.e., the extraction of an energy $-kT \ln 2$ in a cycle. If the demon does not use information from the measurement performing alwyas the same protocol, i.e., $\lambda_t^i = \lambda_t$, then $\tilde{p}_j^i = \tilde{p}_j$ normalized to unity $\sum_i \tilde{p}_i = 1$ yielding $\langle W\rangle = kTD(p_i\|\tilde{p}_i) \geqslant 0$ where $D(p|q)$ is the relative entropy between two probability distributions p and q. To build a Szilard engine, it is enough to find p_i and $\tilde{p}_i^i$ such that the average work $\langle W\rangle$ in Eq. (6.5) is negative; for instance by choosing protocols where $\tilde{p}_i^i > p_i$ (see below for an explicit construction of the engine).

6.3 MEASURING MICROSCOPIC HEAT

Consider a Brownian particle in a potential $U(x,t)$, changing in time, and temperature T. The position of the particle at time t is x_t. Then, the change of potential energy in an interval of time $[t, t+\Delta t]$ is:

$$\Delta U(t) = U(x_{t+\Delta t}, t+\Delta t) - U(x_t, t) \tag{6.6}$$

which can be written as:

$$\Delta U(t) = \delta W(t) + \delta Q(t) \tag{6.7}$$

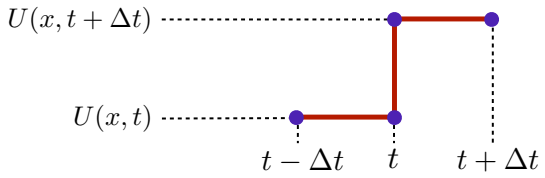
where the work is given by:

$$\delta W(t) = U(x_t, t+\Delta t) - U(x_t, t) \tag{6.8}$$

and the heat:

$$\delta Q(t) = U(x_{t+\Delta t}, t+\Delta t) - U(x_t, t+\Delta t) \tag{6.9}$$

These expressions are exact if the potential changes in discrete steps. Suppose that the potential is constant and equal to $U(x,t)$ between $t-\Delta t$ and t. At time t changes abruptly to $U(x, t+\Delta t)$ and reeq:mainS constant up to $t+\Delta t$ and so on. The change introduces an energy $\delta W(t) = U(x_t, t+\Delta t) - U(x_t, t)$. Then the particle moves from x_t to $x_{t+\Delta t}$ in the constant potential $U(x, t+\Delta t)$. The change in energy $\delta Q(t) = U(x_{t+\Delta t}, t+\Delta t) - U(x_t, t+\Delta t)$ is therefore due to the thermal bath.



If the potential changes smoothly the above expressions are still a good approximation for the work and heat. In fact, they reduce to Sekimoto's expressions for $\Delta t \to 0$:

$$\delta W(t) = \frac{\partial U(x_t, t)}{\partial t}\Delta t \tag{6.10}$$

$$\delta Q(t) = \left.\frac{\partial U(x,t)}{\partial x}\right|_{x=x_t} \circ \Delta x_t \tag{6.11}$$

If $\rho(x,t)$ is the probability density of x_t, the average heat and work are:

$$\begin{aligned} \langle \delta W(t) \rangle &= \int dx\, \rho(x,t)\, [U(x, t+\Delta t) - U(x,t)] \\ &\simeq \int dx\, \rho(x,t) \frac{\partial U(x,t)}{\partial t}\Delta t \end{aligned} \tag{6.12}$$

$$\begin{aligned} \langle \delta Q(t) \rangle &= \int dx\, \rho(x, t+\Delta t) U(x, t+\Delta t) - \int dx\, \rho(x,t) U(x, t+\Delta t) \\ &\simeq \int dx\, \frac{\partial \rho(x,t)}{\partial t} U(x,t) \Delta t \end{aligned} \tag{6.13}$$

The sum of both yields the change of internal energy:

$$\langle \delta W(t) \rangle + \langle \delta Q(t) \rangle = \int dx \, \frac{\partial [\rho(x,t) U(x,t)]}{\partial t} \Delta t \tag{6.14}$$

The total entropy change is:

$$\begin{aligned} \Delta S_{tot}(t) &= -\frac{\langle \delta Q(t) \rangle}{T} - k \int dx \, \frac{\partial [\rho(x,t) \ln \rho(x,t)]}{\partial t} \Delta t & (6.15) \\ &= -\frac{1}{T} \int dx \, \frac{\partial \rho(x,t)}{\partial t} U(x,t) \Delta t - k \int dx \, \frac{\partial \rho(x,t)}{\partial t} \ln \rho(x,t) \Delta t \\ &- k \int dx \, \frac{\partial \rho(x,t)}{\partial t} \Delta t \\ &= -k \int dx \, \frac{\partial \rho(x,t)}{\partial t} \left[\frac{U(x,t)}{kT} + \ln \rho(x,t) \right] \Delta t & (6.16) \end{aligned}$$

If we replace $U(x,t)$ by $-kT \ln \rho(x,t) - kT \ln Z_t$, Z_t being the partition function at time t, then

$$\Delta S_{tot}(t) = k \int dx \, \frac{\partial \rho(x,t)}{\partial t} Z_t \Delta t = 0 \tag{6.17}$$

which is identically zero because $\rho(x,t)$ is normalized at any time t. Notice that, with the replacement $U(x,t) \to -kT \ln \rho(x,t) - kT \ln Z_t$, the total entropy production is zero even far from the quasistatic limit.

6.3.1 *Symmetry breaking*

We analyze how the previous relations change when there is a symmetry breaking along the process. The bead adopts, at the end of the process, one of the two possibilities M and F (for mobile and fixed trap, respectively). The total probability distribution can be written as:

$$\rho(x,t) = p_M \rho_M(x,t) + p_F \rho_F(x,t) \tag{6.18}$$

where p_i is the probability to be in trap $i = M, F$ and $\rho_i(x,t)$ is the probability distribution of x_t conditioned to be at the end of the process in trap $i = M, F$. The expression for the total entropy production conditioned to the final trap can be obtained by repiting the previous arguments:

$$\Delta S_{tot,i}(t) = -k \int dx \, \frac{\partial \rho_i(x,t)}{\partial t} \left[\frac{U(x,t)}{kT} + \ln \rho_i(x,t) \right] \Delta t \tag{6.19}$$

With the replacement $U(x,t) \to -kT \ln \rho(x,t) - kT \ln Z_t$, the total entropy production reads

$$\Delta S_{tot,i}(t) = -k \int dx \, \frac{\partial \rho_i(x,t)}{\partial t} \ln \frac{\rho_i(x,t)}{\rho(x,t)} \Delta t \tag{6.20}$$

We will assume the following scenario for the symmetry breaking:

$$\text{For } t \leqslant t_1: \qquad \rho_M(x,t) = \rho_F(x,t) = \rho(x,t) \tag{6.21}$$

$$\text{For } t \geqslant t_2: \qquad \operatorname{supp}\rho_M(x,t) \cap \operatorname{supp}\rho_F(x,t) = \emptyset \tag{6.22}$$

i.e., the SB is located in a time interval $[t_1, t_2]$, before which the two probability distributions ρ_M and ρ_F are identical. After t_2 the two probability distributions have non overlapping supports. The total entropy production along the whole process $t \in [0, \tau]$ is obtained by integrating (6.20) over time. From $t = 0$ to $t = t_1$, $\rho_i = \rho$ and therefore the integrand is zero. From $t = t_2$ to τ the integral over x is limited to the support of $\rho_i(x,t)$, where $\rho(x,t)/\rho_i(x,t) = p_i$, and therefore:

$$-k\int dx\, \frac{\partial \rho_i(x,t)}{\partial t} \ln \frac{\rho_i(x,t)}{\rho(x,t)} = k \ln p_i \int dx\, \frac{\partial \rho_i(x,t)}{\partial t} = 0 \tag{6.23}$$

In the interval $[t_1, t_2]$ after integration by parts:

$$S_{tot,i} = -k\int_{t_1}^{t_2} dt \int dx\, \frac{\partial \rho_i(x,t)}{\partial t} \ln \frac{\rho_i(x,t)}{\rho(x,t)} \tag{6.24}$$

$$= k \ln p_i + k\int_{t_1}^{t_2} dt \int dx\, \rho_i(x,t) \frac{\partial}{\partial t} \ln \frac{\rho_i(x,t)}{\rho(x,t)} \tag{6.25}$$

$$= k \ln p_i - k\int_{t_1}^{t_2} dt \int dx\, \rho_i(x,t) \frac{\partial}{\partial t} \ln \rho(x,t) \tag{6.26}$$

where we have used $\ln[\rho_i(x,t_2)/\rho(x,t_2)] = -\ln p_i$ and $\ln[\rho_i(x,t_1)/\rho(x,t_1)] = 0$ and the fact that $\int dx\, \rho_i(x,t) \frac{\partial}{\partial t} \ln \rho_i(x,t) = 0$ for all t due to the normalization of $\rho_i(x,t)$.

If the SB occurs in an interval $[t_1, t_2]$ where $\rho(x,t)$ does not change, then $S_{tot,i} = k \ln p_i$. Notice that the replacement $U(x,t) \to -kT \ln \rho(x,t) - kT \ln Z_t$ does not ensure now that the total entropy production vanishes

6.4 EXPERIMENTAL SETUP

We now discuss the experimental setup that we designed to test the universal relations concerning the energetics of the SB and SR. The experiment was done by Ignacio A. Martínez and Dmitry Petrov in *ICFO*, Institute of Photonic Sciences [1].

Figure 52 shows a schematic of the optical apparatus. An acousto-optical deflector (AOD) (ISOMET LS55 NIR) steers a 1060 nm optical beam from a laser coupled into a single-mode fiber (ManLight, ML10-CW-P-OEM/TKS-OTS, maximal power 3 W). The AOD modulation voltage is obtained from an arbitrary

1 Full address: ICFO - Institut de Ciencies Fotóniques, Optical Tweezers Group, Mediterranean Technology Park, Av. Carl Friedrich Gauss, num. 3, 08860 Castelldefels (Barcelona), Spain.

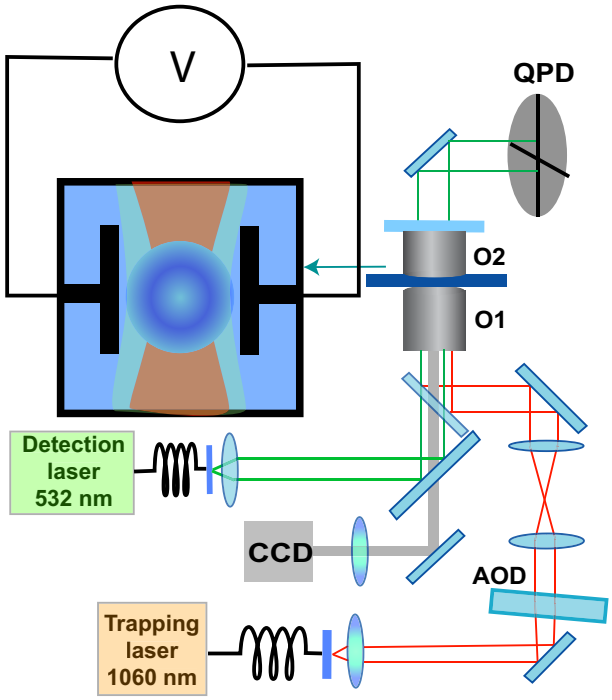


Figure 46: Schematic of the experimental setup used to test the energetics of the symmetry breaking and symmetry restoration.

waveform generator (TaborElectronics WW1071) controlled by a LabView program. The beam deflected by the AOD is expanded and inserted through an oil-immersed objective O1 (Nikon, CFI PL FL 100X NA 1.30) into a custom made fluid chamber. An additional 532 nm optical beam from a laser coupled to a single-mode fiber (OZOptics) is collimated by a $(\times 10,\ NA = 0.10)$ microscope objective and passes through the trapping objective. The forward scattered detection beam is collected by a $(\times 10,\ NA = 0.10)$ microscope objective O2, and its back focal-plane field distribution is analyzed by a quadrant position detector (QPD) (New Focus 2911) at an acquisition rate of 20 kHz. A 532 nm band pass filter in front of the QPD blocks beams with wavelengths different from the detection beam wavelength. The AOD permits the control of the position of the beam focus, and we used in the time-sharing regime to create the protocols described in the text. We used it also for the calibration procedures and for mapping the force distribution in the optical trap. A fluid chamber with microspheres was placed on a piezoelectric-controlled calibrated stage (PiezosystemJena, Tritor 102) allowing 3D translation.

Polystyrene micro spheres (G. Kisker-Products for Biotechnology, http://www.kisker-biotech.com/, polystyrene microparticles PPs-1.0 with a diameter of $1.00 \pm 0.05\,\mu m$) were diluted in distilled de-ionized water to a final concentration of a few spheres per ml. The spheres were inserted into a custom made electrophoretic fluid chamber with two electrodes connected to

a computer controlled electric generator and an amplifier. The chamber was previously described in [Beu12, Mes11].

The individual traps (the fixed trap F, and the other one M that is moved along the x-axis) are not generated simultaneously but with a single optical beam that is rapidly switched between a number of optical focuses by changing the frequency of acoustic waves propagating in the AOD. Multiple optical traps are possible as long as the focused beam is returned to the same location faster than the time it takes for the probe to diffuse away from that location, typically on the order of tens of milliseconds. The time sharing regime of AOD has the greatest additional flexibility because it allows independent control of both the positions of two optical traps (changing the frequency of acoustic waves) and their stiffnesses. In fact, since the probe observes an average intensity, the trapping force at each location may be changed by a proper choice of the timing ratio between two frequency components at the AOD input. Time-sharing techniques using AODs have been described previously (*i.e.*, in [Vis96]).

The alternation of trap positions is controlled by timing signals generated by a modulation generator adjusted to give a low frequency square wave with a controllable duty ratio. The analogue voltage is transferred to the modulation input of a RF driver. The drivers' RF output is then transferred to the AOD. The relative optical intensity in each trap depends on the duty ratio of the modulation signal, and the distance between the traps is defined by the amplitude of the square wave. In order to provide an additional flexibility to the optical trapping system a DC modulation signal was added to the square modulation wave. It permits to shift the positions of both traps simultaneously. We used this signal by a calibration of the traps as it is described below. The time-sharing protocol permitted to control the velocity of the M trap and and the position of its center following prescribed experimental protocols.

The system needs the calibration of the position detection system. At a given position of the quadrant position detector (QPD), the displacements of the sphere produce a unique and linear QPD response only within a restricted region of the probe displacement, usually several hundred nanometers. As we expect an increase of the amplitude of the displacements on application of an additional noise, the linearity range of the detection system and the force map of the optical trap should be characterized.

First, we calibrated the QPD measuring the ratio between the QPD output signals (in volts) and the real displacements of the probe (in nm) by using the natural thermal fluctuations of the probe as described in [Vis96]. The trap was calibrated in the absence of the additional force. Using a relatively strong trap where the Brownian fluctuations did not surpass several tens

of nanometers, we first obtained the PSD of the position of the particle. The experimental PSD is then fitted to a Lorentzian function [Vis96]. From the fit, we obtain the stiffness κ and the calibration coefficient $S_{QPD}(nm/V)$ relating the QPD output signal with the absolute displacement of the sphere.

The spheres were trapped above the bottom surface of the fluid chamber at the distance about 20 μm controlled by the piezoelectric stage. The volume value of the viscous friction coefficient γ was corrected due to the proximity of the chamber surface using the Faxen's law as described in [Vis96].

We then calibrated the acousto-optical deflector (AOD), measuring the ratio between the amplitude of the AOD input modulation signal V_{AOD} (in volts) and the corresponding sphere displacement (in nm) using the previously measured QPD calibration factor S_{QPD}. We moved the trapped sphere in the range of several tens of nanometers, changing the AOD modulation signal, and then measured the output signal of the QPD. This provided us with the calibration factor $S_{AOD}(nm/V_{AOD})$ for the small range of the sphere displacement. We then confirmed with an image analysis that this factor holds for the AOD modulation amplitudes corresponding to the sphere displacements of ± 3 μm.

Figure 47a) shows the dependence of the QPD output signal on the shift of the sphere achieved by changing the AOD modulation voltage. The linear relationship between the displacements of the optically trapped sphere and the output signal of the position detection system can be seen within the range ± 1000 nm. The linear range of the position detection can be extended further using a method suggested recently in [Mar12a].

For completion, we measured the map of the forces that the optical trap exerted on the sphere. These measurements define a range where the linear relationship between the optical force and the sphere displacement is valid. The force map was characterized by the method proposed in [Jah11]. In that method, a strong, stationary, and previously calibrated trap holds a sphere of a given diameter. The optical trapping potential of this trap is strong enough that external forces can not move the sphere out of the range of the harmonic approximation. The intensity of the second trap, whose optical trapping potential for the given sphere has to be analyzed, is adjusted to be low. Its optical force acts as a perturbation exerted on the probe. When the second trap is scanning near the strong trap, it produces a variable force measured through the displacement of the probe in the strong trap.

To experimentally realize this technique, we used the AOD in a time-sharing regime [Mar12b]. Figure 47b) illustrates the force map of the optical trap used in the experiments. The linear range of the restoring optical force extends up to ± 500 nm.

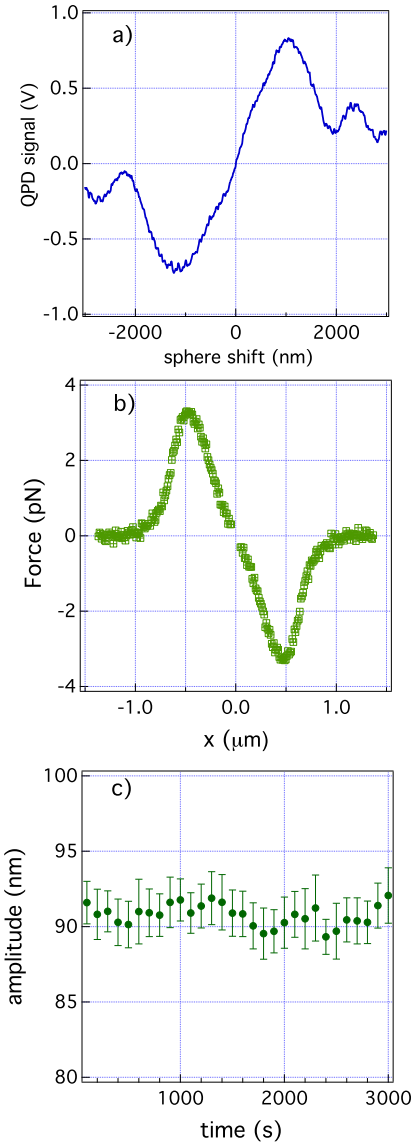


Figure 47: Calibration of our experimental setup. **a**. Output signal of the QPD as a function of the position of the micro sphere. The output signal of the QPD depends linearly on the sphere position when the sphere displacements is within the range $\pm$1000nm. The minor deviations from the linear dependence in the linear part of the QPD response are due to the non uniformity in the detection beam intensity as well as because of the fluctuations of the position of the sphere. **b**. The force exerted on the 1μm optically-trapped sphere versus the radial distance from the trapping beam focus. The incident intensity of the trapping beam at the input pupil of the trapping objective is 5mW. We confirmed that the force map is not affected by further increase of the strong trap stiffness (keeping the same intensity of the weak trap). **c**. The amplitude of the sphere displacement in an AC electric voltage 30V at 1Hz during 3000s.

The calibration curves permitted us to calculate the position of the center of the movable trap and the position of the trapped sphere in the absolute values as presented in Fig. 48.

To characterize the stability of the electric charges on a sphere, we applied a 1 Hz sinusoidal voltage to the electrodes of the electrophoretic chamber; this is within the frequency range of the white noise used in this study. One can also expect changes in the electrical conditions at low frequencies due to possible electrochemical reactions. The amplitude of the displacement of the optically trapped sphere measured with the position detection system did not change more than 6% over at least 3000 s (Fig. 47c)).

6.5 EXPERIMENTAL TEST

Inequalities (6.1) and (6.2) are universal, i.e., they do not depend on the details of the SB or even on the physical nature of the system under consideration. We have tested both inequalities experimentally using a Brownian particle in the potential created by two optical traps.

In the experiment, we trap a polystyrene spherical bead (1 μm diameter) in water with two optical tweezers created by a single laser beam in a time-sharing regime (see Methods). One of the traps, labelled F, is held fixed and the other one, labelled M, is moved along the x−axis following the four step protocol depicted in Fig. 48b: 1) the two traps are initially at rest during a period of time $\tau_1 = 0.5\,s$ with their centers separated by a distance $L_{ini} = 910\,nm$; 2) trap M is moved along the x−axis at constant velocity v_{trap} during a time τ_2; 3) the two traps are again kept fixed for a time $\tau_3 = 0.5\,s$ with their centers separated by $L_{fin} = 1,110\,nm$; 4) trap M is moved back from L_{fin} to its initial position L_{ini} with velocity $-v_{trap}$ in a time τ_2. The total duration of the cycle is $\tau = 2\tau_2 + 1\,s$. By repeating this protocol cyclically, we can study both the SB (steps $1-2-3$) and the symmetry restoration (steps $3-4-1$). Moreover, a uniform electrostatic field created by electrodes in the chamber allows us to bias the motion of the bead towards the M or F trap. The protocol can be considered quasi-static for velocities around 100 nm/s or below, for which heat dissipation due to drag is of order $\gamma v_{trap}^2 \approx 10^{-22}\,J/s \approx 0.02\,kT/s$, where $\gamma = 6\pi R\eta$ is the drag coefficient, $R = 0.5\,\mu m$ being the radius of the bead and $\eta = 8.9 \times 10^{-4}\,Pa \cdot s$ the dynamic viscosity of water at 25°C. We have implemented two quasistatic protocols with $v_{trap} = 100\,nm/s$, $\tau_2 = 2\,s$, and $v_{trap} = 36.36\,nm/s$, and $\tau_2 = 5.5\,s$. In our experiments, the whole cycle is repeated under the same conditions (bead and electrostatic field) for 2400 s, i.e., 480 times for the fast and 200 times for the slow protocol. At the end of the SB protocol (steps $1-2-3$), Kramers transitions

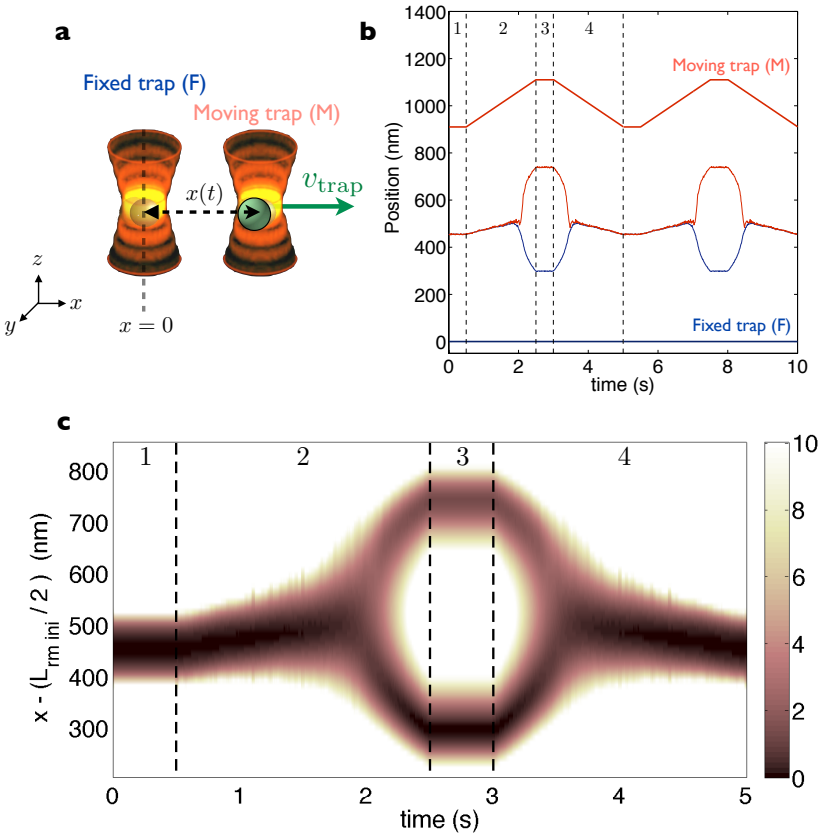


Figure 48: Experimental protocol of symmetry breaking and symmetry restoration. **a**. scheme of the experiment: a trap is held fixed with its center in $x = 0$ and a second trap is moved at constant velocity v_{trap} along the x−axis. The position of the bead $x(t)$ is monitored with sub-nanometer precision. **b**. Position of centers of the two optical tweezers as a function of time during the experimental protocol. The center of the fixed trap F (blue dotted line) stays in $x = 0$ and the center of the moving trap M moves as indicated in red dotted line. The symmetry breaking experiment can be implemented with the cycle formed by the subprocesses $1 \to 2 \to 3$. The average position of the bead after implementing the protocol cyclically during $t = 2,400$ is indicated in blue when the bead chooses the F trap and in red when it chooses the M trap. **c**. Empirical potential $U(x,t) = -\ln \rho(x,t)$ measured with $\rho(x,t)$ obtained from the statistics of trajectories of the bead during $t = 2400$s.

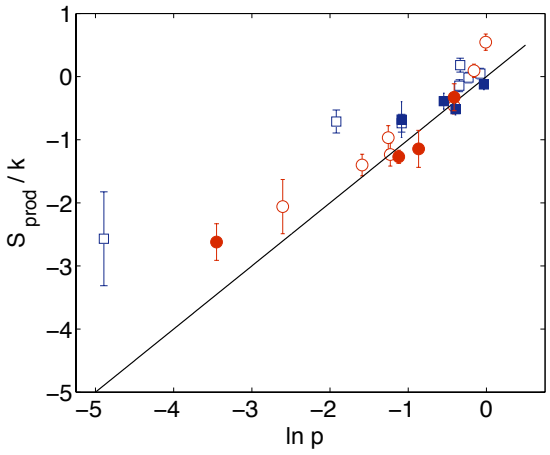


Figure 49: Entropy production in the symmetry breaking as a function of the probability p_i of adopting instance i. Open symbols are obtained using the fast protocol ($\tau_2 = 2\,s$) and filled symbols using the slow protocol ($\tau_2 = 5.5\,s$). Blue squares represent the ensemble averages of ΔS_{prod} for trajectories in which the bead ends in the fixed trap, whereas red circles are the corresponding values for realizations in which the bead ends in the moving trap. Error bars are only statistical and obtained using a statistical significance of 90%.

are not observed and one can unambiguously distinguish two different final meso-states for the bead. depending on the final state of the bead (M or F). We are able to track the position of the bead $X(t)$ with subnanometer precision with an acquisition frequency $f_{acq} = 1kHz$. In Fig. 48b, we plot the average position of the bead of both F and M realizations as a function of time, $\langle X(t)\rangle_{F,M}$, which is calculated as an ensemble average over F and M trajectories, respectively.

An estimation of the potential $V(x,t)$ at each time t is obtained by fitting $-kT \log \rho_{emp}(x,t)$ to a 4-degree polynomial of x with positive leading coefficient. Here $\rho_{emp}(x,t)$ is an empirical probability density function calculated mixing data from both the SB (steps $1-2-3$) and the symmetry restoration part of the process (steps $3-4-1$). Fig. 48c shows the potential obtained for a given realization of the experiment (fast protocol) where we can clearly see the creation of the two wells between 2 and 2.5 seconds.

We have measured the heat or energy transfer to the thermal reservoir (the surrounding water) for individual trajectories (see Methods). The average entropy production up to time t is calculated from the heat and the Shannon entropy, using Eq. (6.2). The entropy production is shown in Fig. 49 as a function of $\ln p_i$ for the SB and $\ln \tilde{p}_i$ for the symmetry restoration. The predicted dependence of the entropy production with respect to the probability is verified in the experiment and the negative entropy production associated to SB transitions is observed as well.

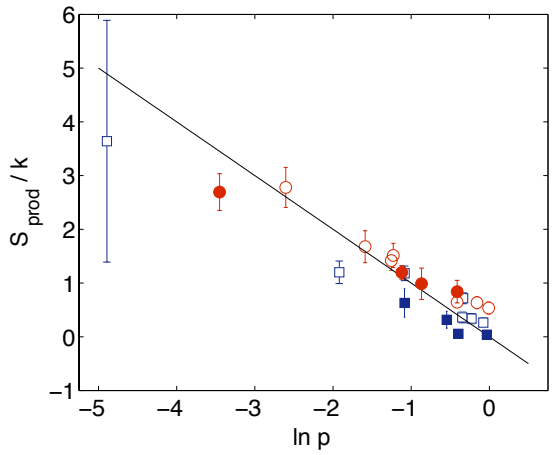


Figure 50: Entropy production in the symmetry restoration as a function of the probability $\tilde{p}_i$ of adopting instance i in the symmetry breaking experiment. Open symbols are obtained using the fast protocol ($\tau_2 = 2\,s$) and filled symbols using the slow protocol ($\tau_2 = 5.5\,s$). Blue squares represent the ensemble averages of ΔS_{prod} for trajectories in which the bead ends in the fixed trap, whereas red circles are the corresponding values for realizations in which the bead ends in the moving trap. Error bars are only statistical and obtained using a statistical significance of 90%.

We notice that we have tested experimentally Eqs. (6.2) and (6.4) rather than their corresponding equations expressed in terms of the dissipative work, Eqs. (6.1) and (6.3). This is due to the fact that we are not able to measure accurately the force exerted by the two traps at any position of the bead. Although the force map of a single trap is known (see Fig. 47b)), experimental observation showed that the potential felt by the bead when the two traps are separated does not coincide with the sum of two equal potentials obtained from the force map. This issue might be solved by using a different experimental setup equipped with a device allowing a direct measurement of the force, and therefore of the work.

6.6 BUILDING A SZILARD ENGINE

As an ilustration of the implications of these results, we construct a Szilard engine that extracts energy from a single thermal reservoir, combining the protocols described above. The engine can be implemented with an adequate combination of SB and symmetry restoration processes where the lower bound for the work $kT \sum_i p_i \ln(p_i/\tilde{p}_i)$ is negative. The minimum is attained for $p_i = 1/2$ and $\tilde{p}_i = 1$, for $i = F, M$, as in the original Szilard cycle. The best we can do in our experiment is $p_F = 0.35$, $p_M = 0.65$, probabilities obtained for an external voltage $V_{ext} = 2\,V$, and

$\tilde{p}_M = 0.99$ for $V_{ext} = 4\,V$ and $\tilde{p}_F = 0.93$ for $V_{ext} = 0\,V$. The Szilard engine consists then on the following feedback protocol: 1) the SB is induced with an external voltage external voltage $V_{ext} = 2\,V$. If the particle end in the fixed trap (blue cufves) we restore the symmetry with no external field, $V_{ext} = 2\,V$. If the particle end in the moving trap (red curves) we restore the symmetry with $V_{ext} = 4\,V$. The cycle should be completed tuning quasistatically the external voltage back to its initial value $V_{ext} = 2\,V$. This last step has not been implemented in the experiment, but it is assumed that it can be realized with arbitrary small entropy production. Fig. 51 shows the average heat (solid curves), the change of Shannon entropy of the probability distribution of the particle (dashed curves) and the average entropy production (dotted curves). The averages are taken over trajectories that end in the fixed trap (upper plot; blue curves), the moving trap (middle plot; red curves), and over all trajectories (lower plot; green curves).

6.7 CONCLUSIONS

The energetics of a generic symmetry breaking (SB) or a symmetry restoration (SR) can be reproduced by the quantitative relationship between irreversibility and dissipation for nonequilibrium processes that occur in the microscopic scale. We have addressed this issue in Chapter 6.1, focusing on the relationship between entropy production and probability to choose an option in a SB.

In chapter 6.2, we have derived an equation that expresses the average entropy production (or energy dissipation) when a microscopic system chooses and instance i — and ends in a subset of the phase space — with the probability to choose that option. Our equation is universal since it does not depend on the mechanism that favors any option i and it is valid for any probability p_i. Our formula correctly predicts that a shrinkage in the accessible phase space leads to a negative entropy production. Moreover, a similar expression is found for a SR, which, together with the equation for the SB, form the core of the thermodynamics of choice. Our two formulas reproduce Landauer principle, the energetics of the Szilard engine and recent expressions obtained in the framework of the generalization of the second law to feedback-controlled processes.

An experimental test of our formulas is done in Chapter 6.5. We have designed an experiment where a Brownian microscopic dielectric particle is trapped with a dual optical tweezer where one of the traps is separated from the other. Using an external DC electric field, we were able to tune the probability of the particle to end in any of the traps, and therefore to implement a SB for different values of p_i. A SR is implemented by retracing the path

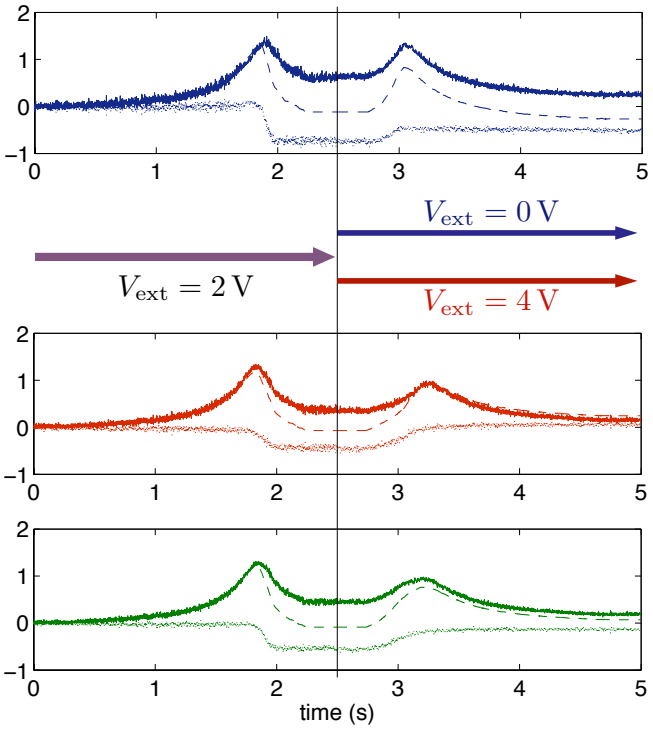


Figure 51: The experimental realization of the Szilard engine. We show the average heat (solid lines) the Shannon entropy of the trajectory distribution (dashed lines), and the average entropy production (dotted lines). The upper plot (blue curves) corresponds to averages over trajectories that end in the fixed trap, the middle plot (red curves) to averages over trajectories that end in the moving trap, and the lower plot shows averages over all trajectories. The feedback protocol is indicated by the arrows. The symmetry breaking is created with an external voltage $V_{\text{ext}} = 2\,\text{V}$ which induces probabilities $p_F = 0.35$, $p_M = 0.65$. When the particle chooses the fixed trap (blue) the symmetry is restored decreasing the voltage to $V_{\text{ext}} = 0\,\text{V}$, and so biasing the potential towards the fixed trap ($\tilde{p}_F = 0.93$). When the particle ends in the moving trap

of the moving trap to its original position. By measuring the average entropy production and the probability to end in the two traps, we reproduce our main formulas derived in Chapter 6.3. The experimental test covers a wide range of probabilities to choose an option, from 5% to 95% .

Our formulas have relevant implications in the thermodynamics of choice of small systems. For example, they can be applied to estimate the probability of a protein to choose between different conformations using the data of the work dissipated along its conformational changes. Moreover, our results suggest that one can design a Szilard engine with the concatenation of an SB with an SR using our setup. In Chapter 6.6, we have explored this possibility by designing such an engine, where the average entropy production after a cycle is negative (but still is positive when taking into account the information entropy).

7

EFFECTIVE HEATING WITH RANDOM FORCES

The use of optical tweezers to study the behavior of microscopic systems driven out of equilibrium has been extended in the last decades, as discussed in Chapter 1.6. Apart from driving Brownian particles under different nonequilibrium scenarios using forces of the amplitude of the fluctuations of the environment, it is also possible to increase the amplitude of the fluctuations by exerting random forces on the particle, as shown experimentally in [GS10]. In the opposite direction, cooling techniques have also been developed and sub-milikelvin kinetic temperatures have been experimentally achieved using Brownian particles recently [Li11]. In this chapter, we show how it is possible to tune experimentally the effective kinetic temperature of a polystyrene microscopic sphere immersed in water by using optical tweezers and random electric fields. Our experimental technique allows one to control the effective temperature of a Brownian particle from room temperature to 3000K (*i.e.* ten times the vaporization temperature of water) in timescales of the order of $\sim$ ms. Our setup can be applied to sample energy landscapes where Brownian particles cannot overcome high energy barriers at room temperature.

When random forces are applied to microscopic systems, the fluctuations of the system, which are affected by both the thermal force and the external random force, are of greater amplitude than if the random external force were not applied. In this situation, the system is said to be in a thermal bath of an *effective* temperature that is higher than the temperature of the surrounding thermal bath. This effective temperature can be increased, for example, using a laser whose wavelength is close to the absorption peak of the environment [Rad09]. Using this technique, one can only increase the effective temperature of a particle up to the vaporization temperature of the material that forms the environment. A different approach, consisting on exerting random forces on a Brownian particle that do not heat the environment, was developed in [GS10]. We follow the latter approach, and achieve temperatures of larger amplitudes than in [GS10] .

Our experiment consists of a Brownian polystyrene spheric bead immersed in water that is trapped with an optical tweezer. The bead is charged, and external random electric fields are applied with two electrodes, which increase the amplitude of the fluctuations of the position of the bead. With our setup, Brownian particles respond to a temperature change in $\sim$ ms

and can achieve a maximum temperature of 3000K without heating the surrounding water. We study both the fluctuations and the response of the bead to nonequilibrium perturbations and analyze the drawbacks that a non-white spectrum of the external random force can cause on the measurements.

This chapter is organized as follows. Section 7.1 is an introduction to the rest of the chapter where we discuss the state of the art of the cooling and heating experimental techniques applied to microscopic systems. In Sec. 7.2 we show the experimental setup that we use to tune the temperature of Brownian particles. Sections 7.3 and 7.4 show our main results concerning the variation of the effective temperature of the particles in an equilibrium and nonequilibrium situations, respectively, and discuss the effect of using correlated external noise instead of pure white noise. In Sec. 7.5 we show how our experiment can be used to accelerate the sampling of the potential created by a dual trap as well as to build microscopic heat engines, and in Sec, 7.6 we discuss all the results of this chapter.

7.1 CONTROLLING THE TEMPERATURE OF SMALL SYSTEMS

Controlling the temperature of small systems is an emerging topic in physics. During the last years, there has been an increasing interest in cooling small systems in order to observe quantum properties that only arise at low temperatures. For example, a cooled optically-trapped system in vacuum can be used to search for non-Newtonian gravity forces at small scales or to produce quantum superpositions of living organisms [Li11] . Since first Ashkin's experiments [Ash77], where the feedback stabilization of optically trapped particle was suggested, cooling of the center-of-mass motion of an optically trapped microspheres has advanced rapidly, and now the cooling below to several millikelvin of microscopic spheres has become possible using active optical feedback [Li11]. This technique has also been applied to nano particles [Gie12]. Such a controllable attenuation of the motion is an important step towards new experiments in different areas of physics. However, we now suggest to go in the opposite direction by controlling the motion of optically trapped microscopic spheres, namely to increase in a controllable way the Brownian fluctuations. We are therefore interested in *heating* instead of cooling the center-of-mass motion of a Brownian particle.

The Brownian motion of microspheres in liquid has offered a model system for experimental studies of processes where thermal fluctuations are important. Out of equilibrium processes in biophysics and colloidal systems [Bus05, Cil10, Bér12], Kramers transitions [McC99], and a micrometer-sized stochastic engine

[Bli11] are just several examples of such processes. We present an experimental technique that allows one to increase the amplitude of the Brownian fluctuations in a controllable way with millisecond response time, extending significantly the range of parameters used in modeling of equilibrium and out-of-equilibrium processes. Our experiment does not require any additional source of heat like optical beams or micro ovens.

Both the temperature of the thermal bath T and its viscous friction coefficient γ affect the amplitude of Brownian fluctuations. We are not aware of methods allowing to change dynamically the viscosity, however, the temperature can be controlled by heating the Brownian sphere or its surroundings. Spherical colloids [Rad09] or nanorods [Rui11] were selectively heated by external sources of a light which is not absorbed by the solvent. In [Mao05, Bli11], the temperature in the surroundings of a dielectric sphere was varied using an optical beam whose wavelength matches an absorption peak of the solvent. The maximum temperature that can be reached using these methods In aqueous experiments is restricted by the the vaporization temperature of the surrounding fluid. In all the mentioned experiments, the temperature increase did not exceed several tenths of kelvin and response times were of the order of several ten of milliseconds.

The fluctuations of the velocity and position are proportional to the kinetic temperature T_{kin}, which for free Brownian particle is equal to the temperature of the surrounding fluid T. Active optical feedback permitted to cool down the kinetic temperature of optically trapped microspheres [Li11, Gie12], and the same technique may increase the kinetic temperature if the active control is tuned to increase the sphere displacements. In this paper we suggest an alternative and simpler method, based on combining an optical trap with a source of additional external forces acting on the trapped sphere. With our technique we achieved the situation when the Brownian sphere fluctuates in water in one of the directions with the center-of-mass motion corresponding to the temperature of several thousand kelvin, however, the viscous coefficient of the liquid and the temperature, describing the center-of-mass motion in two other directions, remain the same as at the temperature of thermal bath.

Consider an overdamped Brownian particle trapped with an optical tweezer of stiffness κ in a fluid at temperature T. For simplicity, we only consider the dynamics in one dimension, x, but the particle can in principle move in three dimensions. The optical trapping potential in $x-$axis, $U(x)$, is assumed to be harmonic and centered at x_0:

$$U(x) = \frac{1}{2}\kappa(x - x_0)^2. \tag{7.1}$$

If an external force $\zeta(t)$ acts on the particle and it is applied in x direction, then the position along the direction of the external force, $x(t)$, obeys the overdamped Langevin equation [cf. Eq.(1.51)]

$$\gamma\dot{x}(t) = -\kappa[x(t) - x_0] + \xi(t) + \zeta(t), \tag{7.2}$$

where γ is the friction coefficient of the Brownian particle in the fluid and $\xi(t)$ is a Gaussian white noise with zero mean and correlation $\langle\xi(t)\xi(t')\rangle = 2kT\gamma\delta(t-t')$, which models thermal fluctuations. We assume that the external force $\zeta(t)$ is also described by a stochastic process independent of the thermal noise, with zero mean and correlation $\langle\zeta(t)\zeta(t')\rangle = \sigma^2\Gamma(t-t')$, where $\Gamma(\tau)$ is a normalized function, $\int_{\mathbb{R}}\Gamma(\tau)\,d\tau = 1$, peaked at $\tau = 0$, and σ is the amplitude of the external noise. If the external noise is white, i.e., $\Gamma(\tau) = \delta(\tau)$, then the particle is subject to an effective noise $\xi_{eff}(t) = \xi(t) + \zeta(t)$, which is a random Gaussian process with zero mean and correlation $\langle\xi_{eff}(t)\xi_{eff}(t')\rangle = 2kT_{kin}\gamma\delta(t-t')$, where T_{kin} is defined as

$$T_{kin} = T + \frac{\sigma^2}{2k\gamma}. \tag{7.3}$$

Hence the position of the particle along fluctuates corresponding to a kinetic temperature T_{kin} which is always higher than the temperature T of the thermal bath. Therefore, a stochastic force exerted on the sphere increases the kinetic temperature of the sphere without real heating the surrounding fluid or the sphere, which remains at temperature T. Notice also that the kinetic temperature along the axes perpendicular to the external force is kept constant and equal to T. The motion of the particle in the direction of the force is in every respect identical to the motion of a Brownian particle immersed in a thermal bath at temperature T_{kin} (this holds for underdamped particles as well). However, this is valid only if the external noise has white power spectral density (PSD). One of the goals is also to describe the limitations arising from the finite correlation time of the external noise, that are unavoidable by experimental realizations when the external force $\zeta(t)$ has a constant PSD only within some frequency range.

7.2 EXPERIMENTAL SETUP

We now discuss the experimental setup that we designed to tune the effective temperature of Brownian microscopic spheres. The experiment was done by Ignacio A. Martínez and Dmitry Petrov in *ICFO*, Institute of Photonic Sciences [1]. The experimental setup

1 Full address: ICFO - Institut de Ciencies Fotóniques, Optical Tweezers Group, Mediterranean Technology Park, Av. Carl Friedrich Gauss, num. 3, 08860 Castelldefels (Barcelona), Spain.

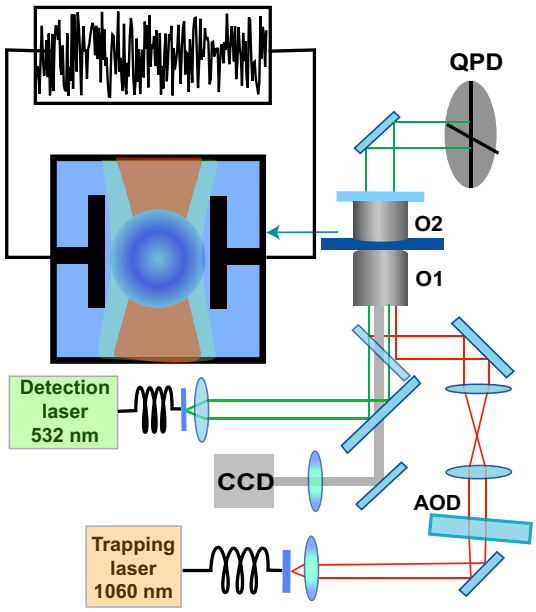


Figure 52: Experimental setup. An acousto-optical deflector (AOD) (ISOMET LS55 NIR) steers a 1060 nm optical beam from a laser coupled into a single-mode fiber (ManLight, ML10-CW-P-OEM/TKS-OTS, maximal power 3 W). The AOD modulation voltage is obtained from an arbitrary waveform generator (TaborElectronics WW1071) controlled by a LabView program. The beam deflected by the AOD, is expanded and inserted through a oil-immersed objective O1 (Nikon, CFI PL FL 100X NA 1.30) into a custom-made fluid chamber. An additional 532 nm optical beam from a laser coupled to a single-mode fiber (OZOptics) is collimated by a $(\times 10, \mathrm{NA} = 0.10)$ microscope objective and passes through the trapping objective. The forward scattered detection beam is collected by a $(\times 10, \mathrm{NA} = 0.10)$ microscope objective O2, and its back focal-plane field distribution is analyzed by a quadrant position detector (QPD) (New Focus 2911) at an acquisition rate of 20 kHz. A 532 nm band pass filter in front of the QPD blocks beams with wavelengths different from the detection beam wavelength. The AOD permits the control of the position of the beam focus. A fluid chamber with microspheres was placed on a piezoelectric-controlled calibrated stage (PiezosystemJena, Tritor 102) allowing the 3D translation. Two electrodes are connected to the two ends of the fluid chamber to exert random forces on the microspheres. The electrodes are connected to an electric generator that can be fed by noisy signals with controllable PSD. The inset of the figure sketches an optically trapped sphere between the electrodes.

is analog to the one used to test our analytical results in the energetics of symmetry breaking, described in Chapter 6.4

In previous experimental works, controllable forces on an optically trapped sphere were applied either with magnetic [Gos02] or optical [Ash77, Li11, Gie12] fields or modulating the position of the trap $x_0(t)$ [GS10]. In our experiment, we apply random electric fields to optically trapped dielectric spheres with electric charges that remain constant, as we show below, during several thousand seconds. The setup includes the following principal parts shown and explained in Fig. 52.

Polystyrene micro-spheres (G.Kisker-Products for Biotechnology, http://www.kisker-biotech.com/, polystyrene microparticles PPs-1.0 with a diameter of $1.00 \pm 0.05\,\mu m$) were diluted in distilled de-ionized water to a final concentration of a few spheres per ml.

The spheres were inserted into a custom-made electrophoretic fluid chamber with two electrodes connected to a computer-controlled electric generator and an amplifier. The chamber was previously described in [Beu12, Mes11]. The spheres were trapped above the bottom surface of the chamber at a distance $\sim 20\,\mu m$ controlled by the piezoelectric stage. The value of the friction coefficient γ was corrected due to the proximity of the chamber surface using the Faxen law as described in [Svo94a].

Electrical signals applied with the electrodes were produced by the generator fed with a noisy sequence close to Gaussian white noise. The sequence was generated using independent distributed random variables $\zeta_{n\delta t}$, with $\langle \zeta_{n\delta t} \rangle = 0$, $\langle \zeta^2_{n\delta t} \rangle = A$, and $\langle \zeta_{n\delta t} \zeta_{m\delta t} \rangle = 0$ for $n \neq m$, A being the variance of the noise [Vol08]. The time interval δt is the clock time of the generator, which is $\delta t = 0.2\,ms$ in the experiments. Notice that δt is a lower bound to the actual correlation time of the external force, which also depends on the response characteristics of the amplifier.

The spheres are trapped using an infrared ($\lambda_{trap} = 1060nm$) laser trap described in Fig. 52. The orientation of the AOD can be controlled externally, which allows one to modify the position of the trap center x_0. A second green detection laser ($\lambda = 532nm$) is used to monitor the position of the particles. A position detection system equipped with a 532nm band pass filter is placed in the back-focal plane of the objective. The position detection is done by a quadrant position detector (QPD), which measures the intensity of the detection laser, and has a $1 - nm$ sensitivity, which is above the resolution of CCD cameras ($\sim 5 - 10nm$). At last, the kinetic temperature of the optically trapped sphere is obtained from the analysis of the output signals of the QPD.

The system needs two calibration procedures previously described in Chapter 6.4: calibration of the position detection system and calibration of the force exerted by the trap on the sphere. The

displacements of the sphere produces a unique and linear QPD response only within a restricted region of the sphere displacement, usually several hundred nanometers. Since an increase of the amplitude of the displacements by application of the additional noise is expected, the linear range of the detection system and the force map of the optical trap need to be characterized.

First, we calibrated the QPD measuring the ratio between QPD output signals (in Volts) and real displacements of the probe (in nm) by using natural thermal fluctuations of the probe as described in [Vis96]. The trap was calibrated in the absence of the additional force. Using a relatively strong trap where the Brownian fluctuations did not overcome several tens of nanometers we first obtained the *Power Spectrum Density* (PSD) of the position of the particle. The PSD is defined as the Fourier transform of the autocorrelation of the position $PSD(f) = \mathcal{F}[\langle x(0)x(t)\rangle]$, f being the frequency. Using the PSD one can not only characterize parameters of the trap such as the stiffness but also detect any misalignment in the trapping setup or the detection system [Vis96]. For a Brownian particle in a harmonic trap and described by an overdamped Langevin equation, the PSD of the position can is a Lorentzian function (see Appendix E.1). In our experiment, we fit the experimental PSD to a Lorentzian function

$$PSD(f) = \frac{1}{2\pi^2\gamma}\frac{kT}{f^2 + f_c^2}, \tag{7.4}$$

where $f_c = \kappa/(2\pi\gamma)$ is the corner frequency. From the fit to (7.4) we obtain the stiffness κ and the calibration coefficient $S_{QPD}(nm/V)$ relating the QPD output signal with the absolute displacement of the sphere as described in [Vis96]. In our experiment, $\kappa = 6pN/\mu m$ and $S_{QPD} = 1280\,nm/V$.

We then calibrated the AOD by measuring the ratio between the amplitude of the AOD input modulation signal V_{AOD} (in Volts) and the corresponding sphere displacement (in nm) using the previously measured QPD calibration factor S_{QPD}. We moved the trapped sphere in the range of several ten nanometers changing the AOD modulation signal and measured the output signal of the QPD. This provided us the calibration factor $S_{AOD}(nm/V_{AOD})$ for the small range of the sphere displacement. We then confirmed using image analysis that this factor holds for the AOD modulation amplitudes corresponding to the sphere displacements within the range $\pm 3\,\mu m$.

Figure 47 **a** shows the value QPD output signal as a function the shift of the sphere achieved by changing the AOD modulation voltage. As it can be seen, there is a linear relationship between the displacements of the optically trapped sphere and the output signal of the position detection system within the range ± 1000 nm. The linear range of the position detection can be extended further using the method suggested recently [Mar12a].

We also measured the force map that the optical trap exerts on the sphere. Our measurements define a range where the relationship between the optical force and the sphere displacement is linear. Within this range, the optical potential is quadratic and Langevin equation (7.2) can be used to describe the dynamics of the microscopic spheres. The force map was characterized by the method first proposed in [Jah11]. In their method, a strong, stationary, and previously calibrated trap holds a sphere of a given diameter. The optical trapping potential of this trap is strong enough, such that external forces can not move the sphere out of the range of the harmonic approximation. The intensity of the second trap, whose optical trapping potential for the given sphere has to be analyzed, is adjusted to be low. Its optical force acts as a perturbation exerting on the probe. When the second trap is scanning near the strong trap, it produces a variable force measured through the displacement of the probe in the strong trap.

For the experimental realization of this technique we used the AOD in a time-sharing regime [Mar12b]. The individual traps are not generated simultaneously as done in [Jah11], but a single optical beam is rapidly switched between a number of optical focuses by changing the frequency of acoustic waves propagating in the AOD. Multiple optical traps are possible as long as the focused beam is returned to the same location faster than the time it takes for the probe to diffuse away from that location, typically on the order of tens of milliseconds. That permits to generate a dual trap with controllable distances between the traps and ratio between their stiffnesses. The alternation of the trap positions is controlled by timing signals generated by a modulation generator adjusted to give a low frequency square wave with a controllable duty ratio. The analogue voltage is transferred to the modulation input of a RF driver whose output is transferred to the AOD. The relative optical intensity in each trap depends on the duty ratio of the modulation signal, and the distance between the traps is defined by the period of the square wave.

Figure 47 **b** shows the force map of the optical trap used in the experiments. The linear range of the restoring optical force extends up to $\pm 500\,\mathrm{nm}$. A slight deviation from the linear dependence for small displacements from the center of the potential has been observed and discussed earlier in [Rico8].

The calibration procedures described above allow to express the experimental data in the physical units for both force exerted by the trap on the particle (pN) and the sphere displacement (nm).

To prove the stability of the electric charges on a sphere, we applied to the electrodes of the electrophoretic chamber a sinusoidal voltage at 1 Hz, i.e. within the frequency range of the external

white noise used in this work. One might expect changes of the electrical conditions at low frequencies due to possible electrochemical reactions. However, the amplitude of the displacement of the optically trapped sphere measured with the position detection system did not changed more than 6% during at least 3000 s as shown in Fig. 47 **c**.

7.3 BROWNIAN SPHERE IN A STATIONARY TRAP WITH ADDITIONAL NOISE

We first investigate how the addition of an external random force alters the dynamics of a Brownian sphere trapped in using an optical tweezer whose center is held fixed. We focus on the analysis on the fluctuations of the position of the particle, whose amplitude increase when adding an external random force. We discuss how one can define an effective kinetic temperature in this situation and how this temperature can be tuned using our setup.

We analyze the time traces of the position of the trapped sphere when the trap center is held fixed, for two cases: (i) in the absence of any external force $\zeta(t) = 0$ and (ii) at the maximum value of the amplitude of the external where the position of the particle does not exceed the linear range of the QPD. PSDs of the position of the sphere are shown in Fig. 53 and Fig. 54 illustrates the probability density of the sphere position using data recordings of $\sim$ 10s. Notice that the data acquisition time exceeds in several orders of magnitude the relaxation time of the particle in the trap $\tau_r \sim$ ms, which is measured from the PSD as we show below. Therefore using a 10s experiment we are sampling the position of the particle after it has reached an stationary state. As shown in Fig. 54, the maximal sphere Brownian excursions without the additional noise do not exceed ± 50 nm, however with the additional stochastic force the maximal amplitude can reach ± 250 nm. Even in this case the sphere displacements do not exceed the linear ranges of both the position detection system and the harmonic approximation of the optical trapping potential holds valid.

Our generator allowed us to obtain an electrical signal with flat spectrum for frequencies up to 1 kHz. As it can be seen in Fig. 53, the PSD in the presence of the additional force is also Lorentzian and the corner frequency f_c does not change, as expected, since it is a function only of κ and γ. We notice that $f_c \sim 100$Hz which implies that the relaxation time of the bead is $\tau_r \sim 1/f_c \sim 10$s On the other hand, the height of the PSD increases due to the external force. Therefore the PSD of the position of the bead when the field is applied is in every respect identical to the PSD of a Brownian sphere trapped by the same tweezer but immersed in a

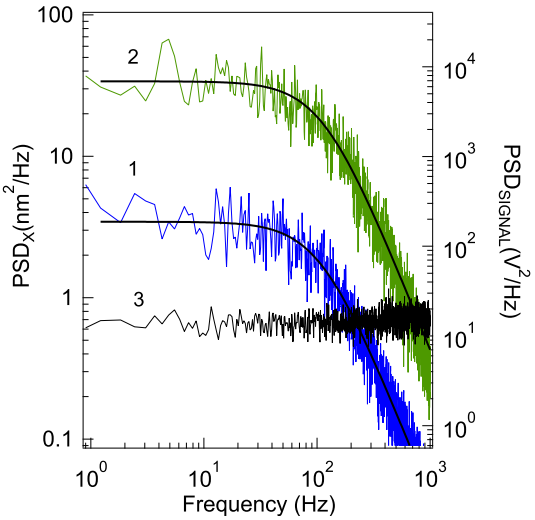


Figure 53: Effect of the addition of a random force on the PSD. Power spectrum density of the position of the sphere obtained from a 10s time series without (blue curve, 1) and with (green curve, 2) the additional stochastic force. Solid black lines correspond to the Lorentzian fits. The power spectral density of the input noisy signal measured at the electrodes of the fluid cell is also shown (black line, right $y-$axis, 3).

thermal bath at a temperature that is larger than the temperature of the water.This temperature increase can also be inferred from the modification of stationary probability density of the position of the sphere, $\rho_{ss}(x)$, depicted in Fig. 54, where it is shown that the additional force broadens the distribution of the position.

We now discuss the results shown in Figs. 53 and 54, which concern the modification of the PSD and the position histograms of the Brownian sphere in a fixed trap when applying an additional random force. As we have seen, the addition of a random external force affects both the auto-correlation and histogram of the position of the bead. Consequently, we can define *two* kinetic temperatures, either using the data of the PSD or $\rho_{ss}(x)$. As we show below, these two kinetic temperatures can differ if the spectrum of the additional noise is not white.

From the Langevin equation (7.2), we can calculate analytically both the PSD and the stationary probability distribution $\rho_{ss}(x)$ for an external noise $\zeta(t)$ with arbitrary correlation $\langle \zeta(t)\zeta(t')\rangle = \sigma^2\Gamma(t-t')$, with $\int_{-\infty}^{\infty}\Gamma(t)dt = 1$ (see details in Appendix E.2). The PSD for such a system is given by:

$$\mathrm{PSD}(f) = \frac{1}{4\pi^2\gamma^2}\,\frac{2\gamma kT + \sigma^2\tilde{\Gamma}(f)}{f^2 + f_c^2}, \tag{7.5}$$

where $\sigma^2\tilde{\Gamma}(f)$ is the Fourier transform of the correlation function of the external noise. If $\tilde{\Gamma}(f)$ is constant for frequencies much higher than the corner frequency of the trap f_c, then the PSD (7.5)

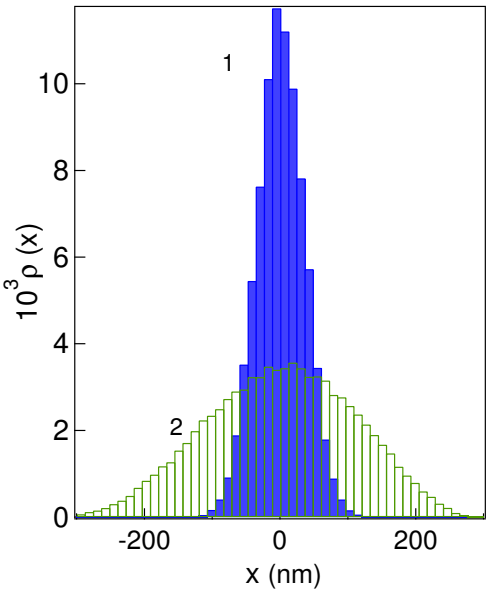


Figure 54: Effect of the addition of a random force on the position histograms. Probability density of the position of the sphere without (1) and with the additional stochastic force (2) corresponding to the PSD curves (1) and (2) shown in Fig. 53.

is approximately a Lorentzian function with the same corner frequency. We can define the following effective kinetic temperature,

$$T_{PSD} \equiv \frac{2\pi^2 \gamma f_c^2 \, PSD(0)}{k} = T + \frac{\sigma^2}{2k\gamma}. \tag{7.6}$$

Notice that, with this definition, T_{PSD} does not depend on the shape of the correlation function of the noise, $\Gamma(t)$.

On the other hand, we can define the following kinetic temperature derived from width of the probability density of the position:

$$T_{hist} \equiv \frac{\kappa \langle x^2 \rangle_{ss}}{k} \tag{7.7}$$

The stationary probability density $\rho_{ss}(x)$ can also be calculated analytically for a Brownian particle described by the Langevin equation (7.2) (see Appendix E.2). The probability density in the stationary state is found to be Gaussian with zero mean and dispersion

$$\langle x^2 \rangle_{ss} = \frac{kT}{\kappa} + \frac{\sigma^2}{\gamma\kappa} \int_0^\infty dt \, e^{-\kappa t/\gamma} \, \Gamma(t), \tag{7.8}$$

yielding

$$T_{hist} = T + \frac{\sigma^2}{\gamma k} \int_0^\infty dt' \, \Gamma(t') e^{-t'/\tau_r}, \tag{7.9}$$

$\tau_r = \gamma/\kappa$ being the relaxation time of the particle in the trap. If $\Gamma(t)$ is peaked around $t = 0$ and the correlation time of the noise, given by $\tau_c \equiv \int_0^\infty dt\, t\, \Gamma(t)$, is small compared to τ_r, T_{hist} is approximatively (see Appendix E.2)

$$T_{hist} \simeq T + \frac{\sigma^2}{2k\gamma}\left(1 - \frac{2\tau_c}{\tau_r}\right). \tag{7.10}$$

Notice that T_{hist}, unlike T_{PSD} defined as (7.6), does depend on the correlation time of the noise. The difference between T_{PSD} and T_{hist} is proportional to τ_c/τ_r, therefore it vanishes if the external noise is a Gaussian white noise. In our experiments, $\tau_r = 1.4$ ms and $\tau_c = 0.26$ ms, being τ_c obtained experimentally from the spectral analysis of the voltage at the electrodes (see below).

Equations (7.6) and (7.9) relate T_{PSD} and T_{hist} with the correlation function of the external noise, respectively. We now analyze if these expressions reproduce the

We now analyze quantitatively how the kinetic temperature of the microscopic sphere changes when we tune the amplitude of the external random force. For this purpose, we measure T_{kin} and T_{PSD} experimentally for different values of the noise intensity. Both effective temperatures increase with the intensity of the external force as shown in Fig. 55. We now test if Eqs. (7.6) and (7.9) relating T_{PSD} and T_{hist} with the correlation function of the external noise, respectively, reproduce the experimental results. First, we obtain T_e for the different experimental data from T_{PSD} asumming that (7.6) is valid. The analytical value of T_{hist} depends not only on the amplitude but also on the correlation function of the external random force, whose shape is now discussed.

In practical situations, Gaussian white noise cannot be implemented and only noisy signals with flat spectrum up to a finite frequency can be designed experimentally. In our experiment, the spectrum of the external random force is flat only up to 1kHz, as shown in Fig. 53. We also measured the autocorrelation function of the signal on the electrodes $V(t)$, defined in [Box11], $\tilde{\Gamma}(t) = \frac{\langle V(t')V(t'+t)\rangle}{\langle V(t')V(t')\rangle}$, which can be fitted to the function $\tilde{\Gamma}(t) = e^{-(t/\tau_c)^2}\cos(t/\tau_o)$, being $\tau_c = 0.26\,$ms and $\tau_o = 0.17\,$ms, as shown in Fig. 56. The correlation of the noise is related to $\tilde{\Gamma}(t)$ by $\Gamma(t) = N\tilde{\Gamma}(t)$, where N is a normalization constant such that $\int_{-\infty}^{\infty}\Gamma(t)dt = N\int_{-\infty}^{\infty}\hat{\Gamma}(t)dt = 1$. The full expression of the correlation function of the additional noise in our experiment is

$$\Gamma(t) = \frac{e^{tc^2/4t_o^2}}{\sqrt{\pi}}\frac{e^{-(t/\tau_c)^2}}{\tau_c}\cos(t/t_o). \tag{7.11}$$

Using (7.11) in (7.9) we calculate analytically the value of T_{hist} as a function of the noise intensity which we show in Fig. 55. In the figure, we plot the effective temperatures T_{PSD} and T_{hist}

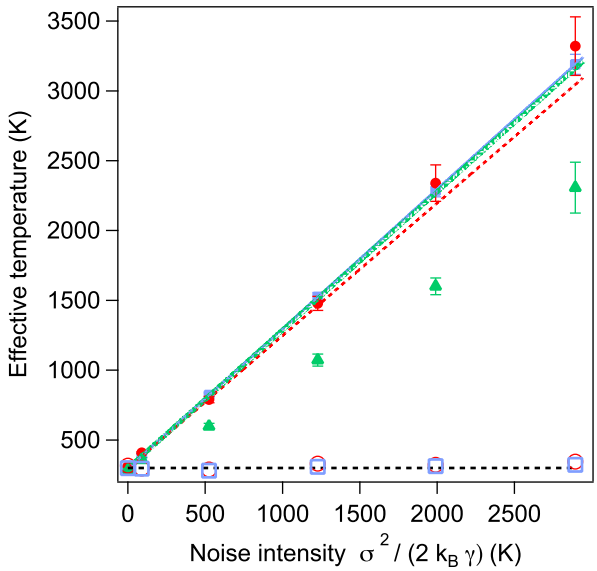


Figure 55: Effective kinetic temperatures as functions of $T_e = \sigma^2/(2k\gamma)$. Filled symbols correspond to experimental values of the effective temperature measured from the fluctuations and the response of the bead along the x axis (direction of the additional noise): T_{PSD} (blue filled squares), T_{hist} (red filled circles) and T_C (green filled triangles). Statistical errors are plotted in the error bars. Lines represent the analytical values obtained for non-white external noise $\zeta(t)$ with correlation function given by Eq. (7.11) for T_{PSD} (7.6) (blue solid line), T_{hist} (7.9) (red dashed line) and T_C (green dotted line), which was obtained analytically using Eq. (7.14). We also show, in open symbols, the experimental values of the effective temperature measured from the fluctuations along the y axis, T_{PSD} (blue open squares) and T_{hist} (red open circles). Statistical errors are smaller than the size of the symbols. Black dashed line is set to 300K.

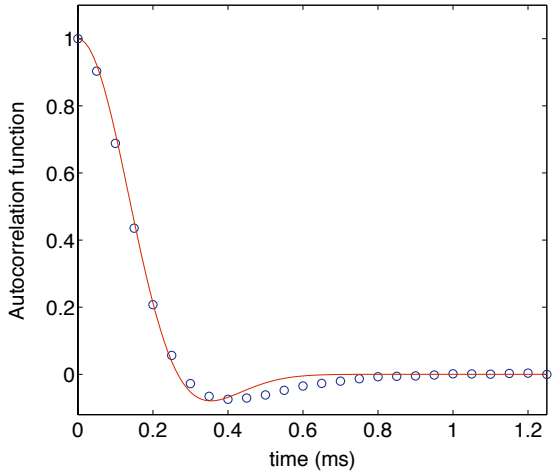


Figure 56: Autocorrelation function of the random force directly measured in the electrodes as a function of time (in milliseconds): experimental data (blue squares) and fit to $\tilde{\Gamma}(t) = e^{-(t/\tau_c)^2}\cos\left(\frac{t}{\tau_o}\right)$ (red curve).

as functions of $T_e = \sigma^2/2k\gamma$, and we observe that T_{hist} and T_{PSD} coincide within statistical errors. For $\Gamma(t)$ defined in Eq. (7.11), we obtain using (7.9) $T_{hist} = T + 0.95\, T_e$ which is slightly smaller than $T_{PSD} = T + T_e$ (7.6) and fits the experimental data.

7.4 A NONEQUILIBRIUM PROCESS

A complete characterization of the effective thermal bath of our experiment might include both measurements from fluctuations and response of the Brownian particle to a nonequilibrium driving. We now study how the presence of the additional random force alters the dynamics of the microscopic sphere in a nonequilibrium experiment. The response of the Brownian particle to an external driving force is related to the temperature, as it was first proved by the fluctuation–dissipation relationship (FDR) (1.43). Since the FDR is only valid when the system is weakly perturbed away from equilibrium, fluctuation theorems could be used to study the response of the system with the advantage that they are valid for any nonequilibrium process. Crooks fluctuation theorem (CFT) establishes a temperature-dependent relationship between the probability density function (PDF) of the work W along a process arbitrarily far from equilibrium, $\rho_F(W)$, and the PDF of the work in the time-reversed process, $\rho_B(W)$ [see Eq. (1.73)]. Notice that CFT goes beyond the fluctuation–dissipation relationship since it is valid not only in the linear but also in non-linear response limit.

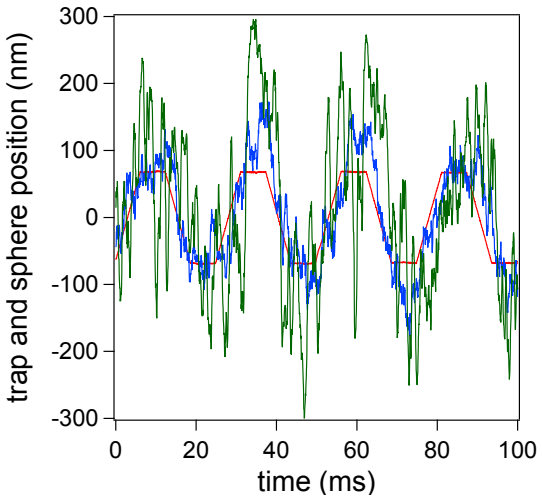


Figure 57: Position of the trap center (black line) and the position of the sphere as a function of time without external electric field (blue curve) and with noisy electric field (green curve).

We study the dynamics of the system under a nonequilibrium driving where the trap center is displaced at a constant velocity. This can be done experimentally by modulate the position of the trap center by feeding the AOD with an appropriate signal.

We have implemented the nonequilibrium protocol depicted in Fig. 57. In the first step, the trap center is moved from $x_0 = -L$ to $x_0 = +L$ at constant velocity $v = 2L/\tau$, where $L = 61\,\mathrm{nm}$ and $\tau = 6.3\,\mathrm{ms}$. This is the forward process. We then let the sphere relax to the new equilibrium position, keeping the trap center at $x_0 = +L$ for the same time $\tau = 6.3\,\mathrm{ms}$, which is larger than the relaxation time of the trap, $\tau_r = \gamma/\kappa = 1.4\,\mathrm{ms}$. Then the trap is moved back from $x_0 = +L$ to $x_0 = -L$ with the same velocity $-v$ (backward process) and it is held fixed at $x_0 = -L$ for the same time τ. This protocol was repeated about 10^4 times. The position of the Brownian particle follows the trap center along this protocol as shown in Fig. 57. We observe that the excursions of the particle out from the trap center significantly increase when adding the external random force.

In each repetition of the nonequilibrium process, we calculate the work in the forward and time-reversed processes using the stochastic trajectories of the position of the bead using Sekimoto's expression for the work associated to a trajectory of an overdamped Brownian particle (1.62). In this particular case, the control parameter is the position of the trap center, $\lambda(t) = x_0(t)$. The work distributions in forward and backward processes $\rho_F(W)$ and $\rho_B(-W)$ are Gaussian as shown in Fig. 58. Moreover, $\rho_F(W)$ and $\rho_B(-W)$ intersect at $W^\star \simeq 0$, in accordance with the prediction of CFT [$W^\star = \Delta F = 0$, cf. Eq. (1.76)]. When adding the

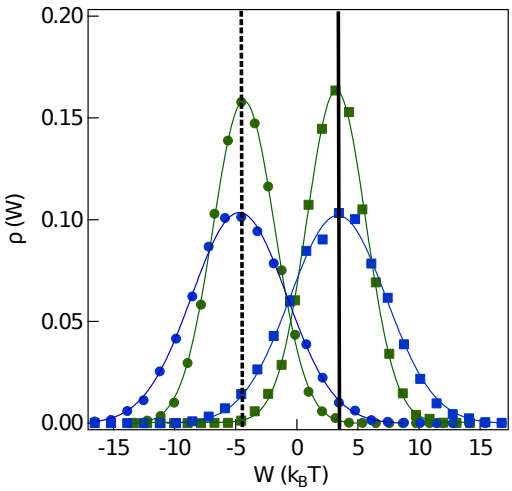


Figure 58: Probability density function of the work obtained in ~ 7000 realizations of the forward process, $\rho_F(W)$ and probability density function obtained from the same number of realizations of the backward process, $\rho_B(-W)$. Squares correspond to the forward process ,$\rho_F(W)$, and the circles to the backward process, $\rho_B(-W)$, where blue and green symbols are obtained without and with additional external force, respectively. Solid blue and green lines are Gaussian fits. Vertical black lines represent the analytical value of the average work in the forward process calculated using Eq. (7.12) (solid line) and the average of $-W$ in the backward process (dashed line) which is calculated in an analogous way to Eq. (7.12).

external noise, both forward and backward work distributions broaden despite their mean value is the same as in the absence of the additional force. The intersection is also $W^\star \simeq 0 = \Delta F$ in this case. We now compare the values of the average work in forward and backward processes with the analytical value derived under the assumption that the particle is described by the overdamped Langevin equation. In Appendix E.3 we show how to calculate the average work done on a Brownian particle by a harmonic trap that moves a constant velocity, assuming overdamped limit. The analytical value of the average work for the forward process is [Jar99]

$$\langle W \rangle = \gamma v^2 \tau \left[1 + \frac{\tau_r}{\tau} (e^{-\tau/\tau_r} - 1) \right]. \tag{7.12}$$

Notice that there is no contribution of the external noise to the average work as long as the external noise enters into Langevin equation as an additive noise. This prediction is reproduced in our experimental results since the average work in both forward and backward processes does not depend on the amplitude of the external noise, as shown in Fig. 58. In addition, the analytical value of the work in the forward process without external field

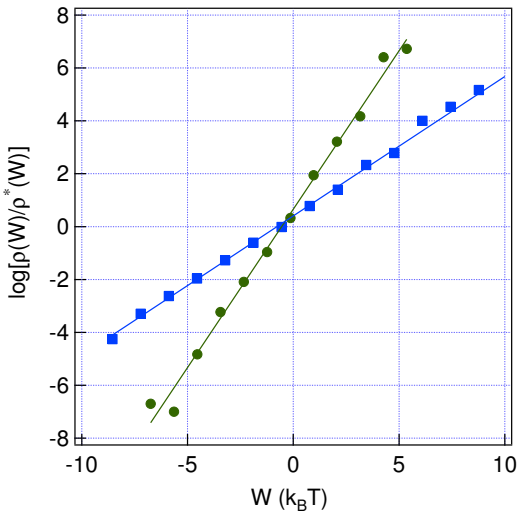


Figure 59: Experimental test of Crooks fluctuation theorem. $\ln[\rho_F(W)/\rho_B(-W)]$ as a function of W/kT without (blue squares) and with additional noise (green circles) with amplitude corresponding respectively to the PSD curves (1) and (2) shown in Fig. 53. The solid lines are linear fits of the experimental data.

is $\langle W\rangle = 14.34\,\text{pN nm}$ whose relative difference with the experimental value, $\langle W\rangle_{\text{exp}} = (15.06 \pm 0.14)\,\text{pN nm}$, is less than 10%. Notice that the actual experimental error of the average work is underestimated since we only take into account the statistical dispersion of the data but not the experimental error. An uncertainty of 0.1pN in the force and 1nm in the yields approximatively an additional error which is of the order of magnitude of the statistical error.

We now investigate the validity of Crooks Fluctuation Theorem (CFT) for different values of the amplitude of the additional noise. Figure 59 shows that $\ln[\rho_F(W)/\rho_B(-W)]$ depends linearly on W with and without additional noise, as expected by the CFT (1.73). In the absence of external noise, CFT is satisfied in our experiment since the slope of $\ln[\rho_F(W)/\rho_B(-W)]$ as a function of W/kT is 1 within experimental errors. With external noise, the slope of $\ln[\rho_F(W)/\rho_B(-W)]$ as a function of W/kT is always smaller than 1 and it decreases when increasing the amplitude of the external noise.

We introduce a new effective temperature, *Crooks Temperature* T_C, to check the consistency of our effective thermal bath for nonequilibrium processes. We define T_C by the inverse of the slope (divided by k) of the linear fit to $\ln[\rho_F(W)/\rho_B(-W)]$ as a function of W, following Eq. (1.73), *i.e.*,

$$\ln\frac{\rho_F(W)}{\rho_B(-W)} = \frac{1}{kT_C}(W - \Delta F). \tag{7.13}$$

Clearly, in the absence of the external force, $T_C = T$, whereas a decreasement of the slope in Fig. 59 when the external noise is added can be interpreted as an *increase* of the effective Crooks temperature. Therefore, nonequilibrium measurements also agree qualitatively with the behavior observed in equilibrium, where the effective kinetic temperature was shown to increase with noise intensity. More precisely, in the absence of external noise, we obtain $T_{hist} = (310 \pm 3)K$, $T_{PSD} = (297 \pm 7)K$ and $T_C = (283 \pm 1)K$, where the errors are purely statistical and therefore a lower bound to the real error. When the external force is added, T_C increases with the noise intensity as shown in Fig. 55 but it lies below T_{PSD} and T_{hist} for every value of the noise intensity. This deviation, which is only observed when the additional force is exerted on the particle, might be caused by the non-white spectrum of the additional force.

We analyze if the deviation between the different effective temperatures observed in our experiment can be explained in terms of the non-whiteness of the spectrum of the additional force. For this purpose, we aim to calculate the value of T_C as a function of the amplitude of the external force, using the actual correlation function of the external force, given by (7.11). The complete analytical calculation of T_C for an overdamped Brownian particle in a driven harmonic trap is shown in Appendix E.4. Since the work distributions are Gaussian, Crooks theorem is given by a simple relationship between the average and the variance of the work,

$$T_C = \frac{\sigma_W^2}{2k\langle W \rangle}. \tag{7.14}$$

In the above equation, $\langle W \rangle$ is given by the expression (7.12). The expression for σ_W^2 is cumbersome and very sensitive to the specific form of the correlation of the external noise, $\Gamma(t)$. T_C can be obtained analytically using (7.14) for a particle described by the Langevin equation (7.2) under an arbitrary external noise with correlation $\Gamma(t)$ and intensity σ^2 as shown in Appendix E.4. Depending on the shape of $\Gamma(t)$, we can get T_C larger or smaller than T_{hist}, for example. The full expression of T_C in terms of the correlation of the additional noise and all the physical parameters of our system can be found in Appendix E.4 in Eq. (E.37). We find that the analytical value of T_C for different noise intensities (dashed line in Fig. 55), is significantly above the experimental value of T_C. We notice that our calculations predict that both for slow driving ($t/\tau_r \gg 1$) and in the white-noise limit ($\tau_c/\tau_r \to 0$) all the temperatures collapse to a single line of slope unity: $T_{PSD} = T_{hist} = T_C = T + T_e$. Both limits were unable to reproduce with our experimental setup because moving the trap slower, the relative error of the work increases and the generator of our setup does not allow to reduce τ_c.

The Lorentzian PSD of the position of the particle when the trap is held fixed is a hallmark of the validity of Langevin equation to describe our system. Therefore we discard that the deviation between T_C and both T_{hist} and T_{PSD} observed in our experiment is due to the finite bandwidth of the additional noise. We also measure the effective temperature from the traces along the y axis, that is, the axis that is orthogonal to the direction of the applied field. To do this, we have previously calibrated the trap in the y axis in the absence of external noise, where we found $\kappa_y = 5.5\text{pN}/\mu\text{m}$ and $S_{QPD,y} = 1700\,\text{nm}/\text{V}$. When increasing the noise intensity, the effective temperature along the y axis remains constant and equal to room temperature (see open symbols in Fig. 55). Therefore, the difference between the experimental and the analytical values of T_C is not due to the presence of an extra degree of freedom where energy is being stored. The presence of torque was also discarded using the technique described in [Volo6] (results not shown). Electrophoretic effects may explain this deviation: in our system, the particle feels a random force which is not exactly the signal recorded from the electrodes due to the reordering of the electric charges of water molecules around the bead. This effect is enhanced when the trap is driven out of equilibrium. Notice that this represents a difference with the experiment introduced in [GS10] where the trap center is moved randomly and Crooks temperature tends to the equilibrium temperature when the correlation time of the external noise tends to zero.

7.5 APPLICATIONS OF OUR SETUP

We now focus on the applications that could be implemented using our setup. Increasing the amplitude of the fluctuations of a Brownian particle might serve to sample energy landscapes where some regions are difficult to access at room temperature. On the other hand, an accurate control of the effective temperature could be used to design microscopic heat engines. We now discuss in detail these two applications.

7.5.1 *Sampling energy landscapes. Kramers transitions.*

Sampling energy landscapes where high energy barriers are present and ergodicity is broken has attracted the interest of many disciplines in physics. In the recent years, single molecule experiments have allowed to explore the energy landscape of macromolecules, showing that the energetic difference between different conformations of RNA [Mao6] or proteins [Onu97, Ceco5] can be of the order of ~ 10 kT. For example, in [Dieo4] Dietz *et al.* measured the energy landscape of a GFP (Green Fluores-

cence Protein) using single molecule techniques, showing that barriers of $\simeq 20$ kT are present in the landscape. In the theoretical framework, different algorithms to sample such kind of energy landscapes have been introduced such as Montecarlo methods [Has70] or umbrella sampling [Tor77].

Let us consider an overdamped Brownian particle in a double-well potential where the two wells are separated by a barrier of height $\Delta E \gg kT$. In equilibrium, the probability of the particle to jump from one well to the other is of the order of $e^{-\Delta E/kT}$ and therefore very small. If an external random force with associated effective temperature $T_{kin} > T$ is applied, the jumping probability increases to $e^{-\Delta E/kT_{kin}} > e^{-\Delta E/kT}$ and the Brownian particle can sample the energy landscape faster than in the absence of the external random force. This simple technique to sample high energy barriers can be implemented using our setup. This situation can be experimentally reproduced by trapping a Brownian particle with a dual optical tweezer. To observe thermally activated escape over a potential barrier, *i.e.* Kramers transitions [Kra40], the optical traps of the dual-well trapping potential have to be close each other, and thermal energy has to be large enough to drive the sphere over the potential barrier between the two neighboring optical traps. The probability distribution of residence times, *i.e.* of the intervals of time between escape events from trap to trap depends on the height of the potential barrier and on the temperature of the system [Kra40]. Experimental results at room temperature are well described by the Kramers theory [Sim92].

We studied how the addition of the external random force affects to Kramers transitions of a microsphere trapped with a dual optical tweezer. We created a double-well potential with controllable distance between two traps and depths of the traps using a time-sharing protocol of the AOD. In Fig. 60a) we show a 10 s−frame of the time traces of the $1\mu m$ diameter sphere moving in the double-well potential before and after the additional noise signal is switched on. When adding the external force, the transitions between the two traps are observed more frequently. In this situation, the potential created by the dual trap can be sampled faster. In Fig. 60b) we show the effective potential generated by the dual trap, obtained as $U(x)/kT = -\ln \rho_{ss}(x)$ using the stationary probability distribution of the position before adding the external force, $\rho_{ss}(x)$. Fig. 60c) shows the residence time probability distribution without and with additional noise signals. As predicted by Kramers [Kra40], the residence time probability decays exponentially with the residence time. Notice that in the presence of the additional noise, the probability distribution decays faster, which is consistent with an increase of the effective temperature.

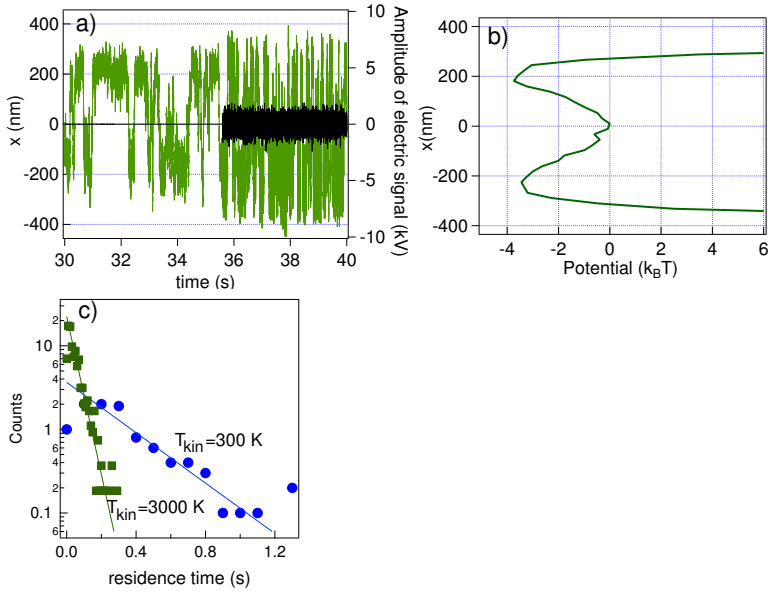


Figure 60: Modification of Kramers transitions with our experimental setup. a) 10−s position traces (green, left axis) of the sphere in the double-well potential as a function of time and voltage on the electrodes (black, right axis) as a function of time. The additional noisy signal was switched on at $t_{switch} = 35.5$ s. b) The trapping potential obtained at room temperature from the stationary distribution of the position before the noise was added. c) Probability of the residence time of the Kramers transitions at room temperature (blue circles) and at 3000K (green squares). The values of the amplitude of the additional force coincides with those used in (1) and (2) cases in Fig. 53.

7.5.2 *Building microscopic heat engines*

Since the origin of thermodynamics, heat engines, *i.e.* devices where a system operates between two thermal baths at different temperature, have been widely studied. The engine of a car, where fuel is in contact with a hot reservoir (the motor) and a cold reservoir (the surrounding air) is a real-life example of a heat engine that performs work. The construction of heat engines in the microscopic scale has been an open problem until recently. Blickle and Bechinger [Bli11] were able to design the experimental version of Stirling engine, where a cyclic process is implemented with two isothermal processes at different temperatures and two isochoric processes. In their experiment, a microscopic sphere is trapped with an optical tweezer. The temperature is controlled using laser heating techniques and the stiffness of the trap plays the role of the volume in original Stirling engine.

Such microscopic engines require an accurate control of the temperature to work in optimal conditions. Our experimental setup is a good candidate to be applied as a temperature control mechanism in a microscopic engine. For this purpose, our setup might be able to control the temperature of the system at fast timescales. We designed an experimental protocol to measure how fast one can switch the effective kinetic temperature of a Brownian particle. We studied the temporal response of the amplitude of the fluctuations of the position of the particle when we change the amplitude of the additional noise abruptly. Fig. 61 shows that the standard deviation of the position of the sphere (and hence the kinetic temperature) changes with the same rate as the electric field amplitude, so that only the relaxation time of the trap ($\tau_r = \gamma/\kappa = 1.4\,\mathrm{ms}$ in our experimental conditions) is the limiting factor. Moreover, the broadening and compressing of the histograms of the position are reversible when we switch on and off the noisy electric field. This fact opens the possibility to implement thermodynamic cycles where a Brownian particle is immersed in a thermal bath of controllable temperature, and therefore to build microscopic heat engines in the spirit of the first microscopic heat engine designed by Blickle and Bechinger [Bli11]. One could implement adiabatic processes by designing a protocol where $\rho(x)$, and therefore entropy of the system along a stochastic trajectory (1.66), $S(t) = -k \int dx\ \rho(x;t)\ \ln \rho(x;t)$, does not change in time. The density $\rho(x)$ can be kept constant in time with our setup by changing the stiffness of the trap and the intensity of the external random force using active feedback.

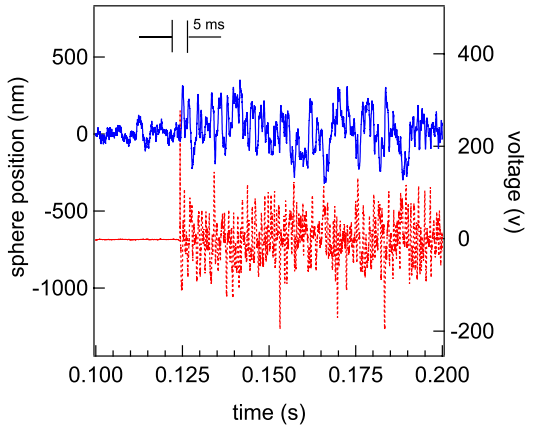


Figure 61: Standard deviation of the position of the sphere (left axis) and amplitude of electric signal (right axis) as a function of time at an abrupt change of the electric field.

7.6 CONCLUSION

In this chapter, we have studied how one can tune the amplitude of the Brownian fluctuations of a microscopic sphere using an experimental setup that is similar to the one used in Chapter 6 to study the energetics of a symmetry breaking. We have studied the dynamics of a Brownian sphere in equilibrium and nonequilibrium situations and discuss the possible applications of our experimental setup.

Our experimental setup was described in Chapter 7.2. We trapped a Brownian microscopic charged sphere in water with an optical tweezer. An external random electric field whose spectrum is flat up to high frequencies is applied to the charged particle and plays the role of an external random force exerted to the microscopic particle. When the optical trap is held fixed and the external field is switched on, both position histograms and PSD suggest that the particle behaves like if it were immersed in a thermal bath at a effective temperature than is higher than room temperature (see Chapter 7.3).

We also studied in Chapter 7.4 the response of the particle to a nonequilibrium process in the presence of the external random field. When the field is switched on, Crooks fluctuation theorem is verified by replacing the room temperature by an effective temperature that is above room temperature. This temperature however lies systematically below the effective temperatures obtained from the PSD of the position histograms with the trap held fixed. We have discarded that this deviation is produced by the non-whiteness of the external noise or torque, and suggested that electrophoretic effects might explain this difference.

Our experiment allows to control the effective temperature of a Brownian sphere within the range 300 K − 3000 K, that is, we can reach effective temperatures that are 10 times greater than the room temperature. The temperature can be changed in a millisecond response time within the range 300 K − 3000 K, which opens the possibility to exploit our technique to design new microscopic heat engines, as discussed in Chapter 7.5. In addition, the setup can be used as a novel experimental technique to sample energy landscapes where ergodicity is broken because of the presence of high-energy barriers.

Part IV

CONCLUSIONS

8

CONCLUSIONS AND OUTLOOK

In this thesis, we have studied the relationship between irreversibility and dissipation for microscopic systems that reach a nonequilibrium steady state (NESS), in the framework of stochastic thermodynamics. We have found a tool to estimate the dissipation of a microscopic system in the NESS from the statistics of a single stationary trajectory produced by the system. Our technique measures the Kullback-Leibler divergence (KLD) between the probability to observe a (*forward*) trajectory and the probability to observe the time reversed (or *backward*) trajectory, and it does not require to know any physical detail of the system to estimate the entropy produced by the system during the process. We have derived analytical expressions for the KLD for some specific stochastic processes and applied our technique to simulations as well as to experimental data acquired from a biological system. A specific nonequilibrium process, namely, a symmetry breaking, is also studied as a relevant example where the quantitative relationship between irreversibility and dissipation holds. We have found universal results concerning the energetics of a symmetry breaking at the microscopic level, and checked these results experimentally using optical tweezers. Our experimental setup has also allowed us to design a protocol to tune the kinetic temperature of Brownian particles. We now outline the main results obtained in this work:

1. **New estimators of the KLD rate.** We have introduced new estimators of the KLD rate between the probability to observe a trajectory and the probability to observed the corresponding time-reversal trajectory. For discrete trajectories, using statistics of blocks or sequences of $m-$data is the most feasible technique to estimate the KLD rate as long as there is sufficient statistics, as described in Chapter 3.1.1. When there is not sufficient statistics, *i.e.* when a finite-data sequence is not sampled in both forward and backward trajectories, the best strategy is to use our estimator based on compression algorithms defined in Chapter 3.1.2, which outperforms previous estimators of the KLD based on compression [Ziv93]. On the other hand, we have also introduced two different techniques to measure the KLD rate for real-valued or continuous time series. First, we have designed a novel estimator that transforms the time series into a directed graph whose in-going and out-going degree distributions capture the irreversibility of a continuous time

series (see Chapter 3.3). This method is applied to several stochastic processes that serve as benchmark cases, showing that it is possible to distinguish between reversible and irreversible stochastic processes as well as between conservative and dissipative chaotic processes. We introduced in Chapter 3.2.2 a second estimator of the KLD rate for continuous correlated time series which uses an asymmetric functional to produce a new uncorrelated series whose KLD is easier to calculate. Such functional is the residual of an auto regressive model. Unlike when using the visibility technique, the KLD estimator using the residual function can be closely related to the average dissipation of a microscopic system in the NESS, which is analyzed using biological data.

2. **Estimating dissipation from single stationary trajectories of discrete systems.** We have first applied our technique to a discrete flashing ratchet model (see Chapter 4). We have compared the average dissipation of the model with the KLD rate between trajectories of the ratchet that contain partial or full information of the system. We have shown that, even when having partial information of the system, our estimators are able to distinguish between equilibrium and NESS. When an external force stalls the ratchet, no net current of particles is observed and linear irreversible thermodynamics fail to predict that the system is dissipating energy. However, using out technique, we correctly predict that the system is producing entropy despite no net current is observed. In this simple example, the KLD, and not the current, is revealed as the actual fingerprint of irreversibility in microscopic systems.

3. **Detecting active processes in biological systems.** In Chapter 5, we have applied our technique to biological (continuous) data. With our KLD rate estimator based on the use of residuals we have been able to distinguish between active and passive spontaneous oscillations of ear hair bundles from bullfrog's sacculus. Our technique only requires the measurement of spontaneous oscillations of the position of the top of the hair bundles during ~ 100s, outperforming significantly the requirements of previous works [Mar01b]. In addition, our KLD rate estimator indicates the minimum irreversibility to sustain spontaneous oscillations in bullfrog's ear hair bundles, which is of the order of the energy consumption rate from ATP hydrolyzation by a single adaptation motor. Therefore our method serves not only to detect the presence of active processes but also to bound the entropy produced by such processes when they are present.

4. **Universal features of the energetics of the symmetry breaking.** The energetics of a symmetry breaking and a symmetry restoration can be linked to the relationship between dissipation and irreversibility in the microscopic scale. When a microscopic system breaks a symmetry, the accessible phase space of the system is reduced to a subset of the full phase space. Even when the symmetry is broken quasi statically, the symmetry breaking is accompanied by a (negative) entropy production. We have derived a formula that relates the average entropy production when the system chooses an instance with the probability of the system to choose that option. Our formula is valid in the quasi static limit and it is universal: it does not depend on the mechanism that biases the probability to choose any option. In addition, we have obtained an equivalent formula for the symmetry restoration process, which together with the symmetry breaking result, can be used to derive Landauer's principle and the energetics of the Szilard engine as discussed in Chapter 6. For a complete validation of our formula, we have checked our results experimentally, by trapping a Brownian particle in a dual optical tweezer and using external electric fields to tune the probability of the bead to choose between the two traps.

5. **Building experimental microscopic thermodynamic engines.** The development of experimental techniques in this thesis has provided us a tool to design a variety of thermodynamic processes in the microscopic scale. Using the dual optical trap, we have been able to reconstruct a Szilard engine as a concatenation of a symmetry breaking and a symmetry restoration, as discussed in Chapter 6. On the other hand, we have also designed an experiment to tune the amplitude of the Brownian fluctuations of a microscopic charged sphere. The setup is discussed in Chapter 7 and consists of a Brownian particle trapped by an optical tweezer and an external random electric field of tunable amplitude. With our setup, we are able to control the kinetic temperature of a microscopic particle from room temperature to 3000K in a fast timescale ($\sim$ ms). This technique opens the possibility of implementing heat engines in the microscopic scale such as a non-quasistatic Carnot engine.

In summary, we have performed a complete study on the relationship between dissipation and irreversibility for microscopic system in the NESS. Our study includes analytical results, simulations and experimental verifications. Our results provide new insights in stochastic thermodynamics and nonequilibrium statistical mechanics but also in time series analysis, information

theory or biophysics. Apart from the aforementioned applications, our work opens the possibility to analyze new problems in stochastic thermodynamics. Here we indicate some of the possible future research lines related to our work

1. **Application of our technique to different biological data.** Once our technique to estimate the dissipation in microscopic systems has been validated using data from bullfrog's ear hair cells, the method could be applied to other biological data. For example, one could analyze if the time asymmetry observed in the heartbeat of an aging heart is related to the dissipation rate of the heart [Cos05]. Also, EEG recordings could be used to estimate the entropy produced in neural processes, or use the KLD to distinguish between different (active or passive) oscillations in red blood cells [Lóp12].

2. **Extension of our method to quantum systems.** All the results of this thesis are in the realm of classical physics and therefore their validity extends to small systems of size where quantum effects are negligible. The quantification of the relationship between irreversibility and dissipation for quantum system has not been considered yet. However, the extension of the KLD to quantum systems has already been discussed in [Ved02].

3. **New experimental insights in stochastic thermodynamics.** Our experimental techniques can be applied to design new microscopic engines such as a version of a Szilard engine simpler than the experiment done in Ref. [Toy10]. The implementation of adiabatic processes in the microscopic scale — and therefore the construction of a Carnot engine — is also possible by using active feedback when tuning the kinetic temperature of a Brownian particle with our experimental setup described in Chapter 7. In addition, the dynamics of trapped Brownian particles under the action of random forces with arbitrary spectrum opens new questions that could be addressed with new experimental observations.

Part V

APPENDIX

A

APPENDIX TO CHAPTER 1

A.1 GAUSSIAN WHITE NOISE

Let us consider the Brownian motion of a particle in one dimension x. We consider the *Wiener process*, $W(t)$, that is the continuous limit of the following discrete Brownian motion in one dimension

$$x(t+\Delta t) = x(t) + \eta_t \Delta x, \tag{A.1}$$

where η_t is a random number that can take two values, $\eta_t = \pm 1$ each of them with probability $p = 1/2$, and Δx is the amplitude of the jumps of the particle. The initial condition is assumed to be $x(0) = 0$. Equation (A.1) models the motion of a Brownian particle in one dimension that starts in $x(0) = 0$ and jumps randomly to the left and to the right with equal probability. The Wiener process $W(t)$ is the limit of this Brownian motion when $\Delta x \to 0$, $\Delta t \to 0$ with $\sigma^2 = \Delta x^2/\Delta t$ finite. The position of the Wiener process satisfies the following properties:

1. $W(t)$ is Gaussian distributed with zero mean and standard deviation $\sigma\sqrt{t}$, $W(t) \sim \mathcal{N}(0, \sigma\sqrt{t})$.
2. $W(t)$ is Markovian, *i.e.* the value of $W(t)$ is independent on the value of $W(t')$ for any $t' < t$.
3. The increment of $W(t)$ is independent on its value, *i.e.* $W(t+\Delta t) - W(t)$ is independent of $W(t)$.

We can investigate which is the value of the correlation of the Wiener process $\langle W(t)W(t')\rangle$ by taking into account the above properties. Because of the 3rd property,

$$\langle [W(t+\Delta t) - W(t)]W(t)\rangle = 0, \tag{A.2}$$

or equivalently

$$\langle W(t+\Delta t)W(t)\rangle = \langle W(t)W(t)\rangle = \sigma^2 t, \tag{A.3}$$

where we have used the 1st property in the second equality. For any t and t',

$$\langle W(t)W(t')\rangle = \sigma^2 \min(t, t'). \tag{A.4}$$

Gaussian white noise is formally defined as the derivative of a Wiener process

$$\xi(t) = \frac{dW(t)}{dt}. \tag{A.5}$$

However, since $W(t)$ is stochastic, its time derivative cannot be computed, and the above definition is not useful for generating a stochastic series of a Gaussian white noise process. However, since $\xi(t)$ is obtained from a Gaussian process, $\xi(t)$ is also a Gaussian random variable. Its average and correlation can be computed from the properties of the Wiener process. Since $\langle W(t)\rangle = 0$,

$$\langle \xi(t)\rangle = \frac{d}{dt}\langle W(t)\rangle = 0. \tag{A.6}$$

The correlation of the white noise can be obtained taking into account (A.4)

$$\langle \xi(t)\xi(t')\rangle = \frac{d}{dt}\frac{d}{dt'}\langle W(t)W(t')\rangle = \frac{d}{dt}\sigma^2\theta(t'-t) = \sigma^2\delta(t-t'). \tag{A.7}$$

Therefore, a Gaussian white noise of amplitude σ is a stochastic process defined by the following three properties

1. It is Gaussian distributed.
2. It has zero mean $\langle \xi(t)\rangle = 0$.
3. Its correlation is such that $\langle \xi(t)\xi(t') = \sigma^2\delta(t-t')$.

A.2 STOCHASTIC DIFFERENTIAL EQUATIONS.

Let us consider stochastic differential equations of the form

$$\dot{x}(t) = f(x(t)) + g(x(t))\xi(t) \tag{A.8}$$

for any generic functions f and g, and $\xi(t)$ being a Gaussian white noise (see Appendix A.1). The solution of these equations have the form

$$x(t) = x(0) + \int_0^t f(x(t'))dt' + \int_0^t g(x(t'))dW(t'). \tag{A.9}$$

Because of the stochastic nature of $W(t)$, the calculation of the last term of (A.9) is cumbersome. We have to give a prescription to the integral

$$\int_0^t g(x(t'))dW(t') = \lim_{\delta t\to 0} \sum_i g(t_i^\star)[W(t_i+\Delta t) - W(t_i)]. \tag{A.10}$$

There are two prescriptions commonly used to calculate this integral introduced by Ito and Stratonovich:

- **Ito**: $t_i^\star = t_i$.
- **Stratonovich**: $t_i^\star = t_i + \frac{\Delta t}{2}$.

Therefore, the integrals of the type $\int g(x(t'))dW(t')$ can be evaluated according to these two different conventions. In general, the following notation [Sek10] is used: For Ito calculus, (A.10) is evaluated using the following discretization scheme,

$$g(x(t'))\cdot dW(t') \to g(x(t'))\left[W(t'+\Delta t) - W(t')\right], \tag{A.11}$$

whereas for Stratonovich calculus

$$g(x(t'))\circ dW(t') \to \frac{g(x(t'+\Delta t)) + g(x(t'))}{2}\left[W(t'+\Delta t) - W(t')\right]. \tag{A.12}$$

Depending on the prescription used, the values of integrals that involve stochastic terms may differ. In the case in which the noise is *additive*, that is, when the term that multiplies $\xi(t)$ does not depend on $x(t)$ ($g(x(t)) = \text{const}$), Ito and Stratonovich calculus yield the same result. In any other case, the two prescriptions may lead to different values. As an example, let us consider the following differential equations

$$\dot{x}(t) = \xi(t), \tag{A.13}$$

$$\dot{y}(t) = x(t)\xi(t). \tag{A.14}$$

The solution to this equation is (see Appendix A.1)

$$y(t) = \int_0^t W(t')dW(t'). \tag{A.15}$$

We now analyze which is the value of the ensemble average of $y(t)$ to different realizations. In Ito formalism,

$$\langle y^{\text{Ito}}(t)\rangle = \lim_{\Delta t\to 0}\sum_i \langle W(t_i)[W(t_i+\Delta t)-W(t_i)]\rangle = 0, \quad \text{(A.16)}$$

where we have used (A.2). On the other hand, in the Stratonovich prescription,

$$\begin{aligned}\langle y^{\text{Strat}}(t)\rangle &= \lim_{\Delta t\to 0}\sum_i \langle W(t_i+\Delta t/2)[W(t_i+\Delta t)-W(t_i)]\rangle \\ &= \lim_{\Delta t\to 0}\sum_i \sigma^2[t_i+\Delta t/2-t_i] = \frac{\sigma^2 t}{2}. \quad \text{(A.17)}\end{aligned}$$

Notice that in Stratonovich calculus usual rules of calculus do apply. In this example,

$$y^{\text{Strat}}(t) = \int_0^t W(t')dW(t') = \frac{W(t')^2}{2}\Big|_0^t = \frac{W(t)^2}{2}. \quad \text{(A.18)}$$

From the above (exact) solution we can calculate the average $\langle y^{\text{Strat}}(t)\rangle = \langle W(t)^2/2\rangle = \sigma^2 t/2$, recovering (A.17). In Ito calculus, usual rules of calculus do not apply. This can be seen intuitively in this example. The term,

$$\langle W(t_i+\Delta t)[W(t_i+\Delta t)-W(t_i)]\rangle = \frac{[W(t_i+\Delta t)^2-W(t_i)^2]}{2} - \frac{[W(t_i+\Delta t)-W(t_i)]^2}{2}. \quad \text{(A.19)}$$

The first term leads to the result from standard calculus $W(t)^2/2$ and the second vanishes if $W(t)$ is differentiable since $[W(t_i+\Delta t)-W(t_i)]^2 \sim \Delta t^2$. However, for the Wiener process, $[W(t_i+\Delta t)-W(t_i)]^2 = \sigma^2\Delta t$, and the contribution of the second term in (A.19) cannot be neglected. In Ito calculus, the terms that go like ΔW^2 are replaced by $\sigma^2\Delta t$. As a consequence, the following property, known as *Ito's lemma*, holds in Ito's convention: If $x(t)$ follows (A.8), and $A = A(x)$ is any function of $x = x(t)$, then

Ito's lemma

$$\dot A(t) = \frac{\partial A}{\partial x}[f(x(t)) + g(x(t))\xi(t)] + \frac{\sigma^2}{2}\frac{\partial^2 A}{\partial x^2}g(x(t))^2. \quad \text{(A.20)}$$

Another paradigmatic case study in which Ito and Stratonovich calculus do not coincide is $\langle x(t)\xi(t)\rangle$. In Ito's formalism

$$\langle x(t)\xi(t)\rangle^{\text{Ito}} = 0, \quad \text{(A.21)}$$

whereas in the Stratonovich prescription, $\langle x(t)\xi(t)\rangle$ does not vanish in general and it is calculated using Novikov's theorem [Nov65b],

$$\langle x(t)\xi(t)\rangle^{\text{Strat}} = \int_0^t ds\langle \xi(t)\xi(s)\rangle \left\langle \frac{\delta x(t)}{\delta \xi(s)}\right\rangle, \quad \text{(A.22)}$$

where $\frac{\delta x(t)}{\delta \xi(s)}$ stands for functional the functional derivative of $x(t)$ with respect to $\xi(s)$.

A.3 FOKKER-PLANCK EQUATION

Let us consider the following generalized overdamped Langevin equation

$$\dot{x} = f(x) + g(x) \cdot \xi, \tag{A.23}$$

where the stochastic product is taken in Ito's sense. By Ito's lemma, any function $A = A(x)$ follows

$$\dot{A} = \frac{\partial A}{\partial x}\left[f(x) + g(x)\xi(t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 A}{\partial x^2}g(x)^2. \tag{A.24}$$

We now focus on calculating the average of $\langle \dot{A} \rangle$. We notice that in Ito's convention $\langle g(x)\xi \rangle = 0$ for any function g (A.21). Therefore,

$$\langle \dot{A}(t) \rangle = \left\langle A'(x)f(x) + \frac{\sigma^2}{2}g(x)^2 A''(x) \right\rangle, \tag{A.25}$$

where we have simplified the notation by using $x = x(t)$ and $A'(x) = \frac{\partial A}{\partial x}$. Any average can be expressed in terms of the probability density function $\rho(x,t)$, $\langle A \rangle = \int dx \rho(x,t) A$, where the integration is done over all the space. Taking into account this, and using the notation $\partial_\alpha = \frac{\partial}{\partial \alpha}$ and $\partial_\alpha^2 = \frac{\partial^2}{\partial \alpha^2}$ for any variable α, we can rewrite (A.25)

$$\int A(x)\frac{\partial}{\partial t}\rho(x,t) = \int dx \rho(x,t)\left[A'(x)f(x) + \frac{\sigma^2}{2}g(x)A''(x)\right]. \tag{A.26}$$

Now, integrating by parts and assuming that $\rho(x,t)$ vanishes in $x = \pm\infty$ at any time t,

$$\int A(x)\frac{\partial}{\partial t}\rho(x,t) = \int dx A(x)\left[-\partial_x(f(x)\rho(x,t)) + \frac{\sigma^2}{2}\partial_x^2(g(x)\rho(x,t))\right]. \tag{A.27}$$

Since Ito's lemma holds for any function $A(x)$, the above equality implies

Fokker-Planck equation

$$\partial_t \rho(x,t) = -\partial_x\left[f(x)\rho(x,t)\right] + \frac{\sigma^2}{2}\partial_x^2\left[g(x)^2\rho(x,t)\right]. \tag{A.28}$$

which is known as the Fokker-Planck equation for the probability density $\rho(x,t)$. Fokker-Planck equation can be seen as a continuity equation for the probability current $j(x,t)$

$$\partial_t \rho(x,t) = -\partial_x j(x,t), \tag{A.29}$$

with

$$j^{\text{Ito}}(x,t) = \left[f(x) - \frac{\sigma^2}{2}\partial_x g(x)^2\right]\rho(x,t). \tag{A.30}$$

In the case of Stratonovich calculus, the definition of the probability current changes slightly,

$$j^{\text{Strat}}(x,t) = \left[f(x) - \frac{\sigma^2}{2} g(x) \partial_x g(x) \right] \rho(x,t). \tag{A.31}$$

For additive noise, both conventions lead to the same result. If $g(x) = C$ with $C = \text{const}$,

$$j^{\text{Ito}}(x,t) = j^{\text{Strat}}(x,t) = \left[f(x) - \frac{\sigma^2}{2} C^2 \right] \rho(x,t). \tag{A.32}$$

B

APPENDIX TO CHAPTER 2

B.1 THE KULLBACK-LEIBLER DIVERGENCE AND ITS PROPERTIES

The Kullback-Leibler divergence or relative entropy between two probability distributions of a random continuous variable X is defined by

$$D[p(x)||q(x)] = \int_{\mathcal{X}} dx\, p(x) \ln \frac{p(x)}{q(x)} \tag{B.1}$$

where the integral runs over all the values that the value X can take, $\mathcal{X}$. We use the conventions $0 \ln \frac{0}{0} = 0$, $0 \ln \frac{0}{q} = 0$ and $p \ln \frac{p}{0} = \infty$. The latter case implies that if there is a value x of the variable X such that $p(x) \neq 0$ and $q(x) = 0$, the KLD diverges. The relative entropy $D[p(x)||q(x)]$ is only finite if the support set of p is contained in the support set of q.

Notice that the definition of KLD can be extended to distribution of any number n of random variables. We define by $X_1, X_2, \cdots, X_n$ a set of n different random variables, which can be simplified using the notation $X_1^n \equiv X_1, \cdots, X_n$. In this case, we denote by $p(x_1^n)dx_1^n = p(x_1, \cdots, x_n)dx_1 \cdots dx_n$ the probability to observe the n random variable to take the values $[x_1, x_1 + dx_1], \cdots, [x_n, x_n + dx_n]$. The KLD between two probability distributions of the set of random variables X_1^n is defined as

$$D[p(x_1^n)||q(x_1^n)] = \int_{\mathcal{X}_1^n} dx_1^n\, p(x_1^n) \ln \frac{p(x_1^n)}{q(x_1^n)}. \tag{B.2}$$

In case of a single random variable of many dimensions such as $\vec{x} = (x, y, z)$, the KLD between two distributions of such a variable can be expressed using (B.2) as the KLD between the joint probability distributions of the components, $p(x, y, z)$ and $q(x, y, z)$ in this example. We now introduce some of the most relevant properties of the KLD that are valid for both single and multivariate probability density functions of discrete and continuous random variables.

We now consider the discrete version of (B.1) to prove the main properties of the KLD [Cov06]. For a discrete random variable X which can take values in a set $X = \{x \in \mathcal{X}\}$, the KLD between two distributions of the variable X, p and q is

$$D[p(x)||q(x)] = \sum_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}. \tag{B.3}$$

We now show and prove some properties of the KLD between two probability distributions of discrete variables that are extensive to the KLD between distributions of continuous random variables as well as to the KLD between probability distributions of more than one variable.

Positivity: For any p and q,

$$D[p(x)||q(x)] \geqslant 0, \tag{B.4}$$

with equality if and only if $p = q$ for all x.

Proof. Let us consider $-D[p(x)||q(x)]$ where we denote by $\mathcal{X}$ the support of X and by A the support set of p, which is contained in $\mathcal{X}$, $A = \{x|p(x) > 0\}$. The following inequalities hold

$$\begin{aligned} -D[p(x)||q(x)] &= \sum_{x\in A} p(x) \ln \frac{q(x)}{p(x)}, && \text{(B.5)} \\ &\leqslant \ln \sum_{x\in A} p(x)\frac{q(x)}{p(x)} = \ln \sum_{x\in A} q(x) && \text{(B.6)} \\ &\leqslant \ln \sum_{x\in \mathcal{X}} q(x) = \ln 1 = 0 && \text{(B.7)} \end{aligned}$$

Since $-D[p(x)||q(x)] \leqslant 0 \Rightarrow D[p(x)||q(x)] \geqslant 0$. WHEN EQUAL SIGN?? Notice that the inequality in (B.6) is consequence of Jensen's inequality for the (concave) function $\ln(x)$, which satisfies $\langle \ln x \rangle \leqslant \ln\langle x \rangle$ for any value of the argument x.

Chain rule: Kullback-Leibler divergence decreases upon coarse graining,

$$D[p(x,y)||q(x,y)] \geqslant D[p(x)||q(x)]. \tag{B.8}$$

Proof

$$\begin{aligned} D[p(x,y)||q(x,y)] &= \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{q(x,y)}, \\ &= \sum_{x,y} p(x,y) \ln \frac{p(x)p(y|x)}{q(x)q(y|x)}, \\ &= \sum_{x,y} p(x,y) \ln \frac{p(x)}{q(x)} + \sum_{x,y} p(x,y) \ln \frac{p(y|x)}{q(y|x)}, \\ &= D[p(x)||q(x)] + D[p(y|x)||q(y|x)], \\ &\geqslant D[p(x)||q(x)]. && \text{(B.9)} \end{aligned}$$

The inequality in (B.9) is consequence of the relative $D[p(y|x)||q(y|x)]$ being positive, which can be proved in an analogous way as $D[p(x)||q(x)] \geqslant 0$.

The properties that are mentioned above apply not only to distributions of independent random variables but also to *dependent*

random variables. A paradigmatic case of dependent random variables is to consider the outcomes of a stochastic process with memory. The KLD can be extended to distributions of stochastic trajectories of a random variable X. The extension is done in a similar way as in the definition of the KLD between probability distributions of n different random variables. Let us consider a finite sequence of n symbols $x_1, x_2, \cdots, x_n$ which we denote by x_1^n. The extension is done. Let $p(x_1^n) = p(x_1, x_2, \cdots, x_n)$ be a joint probability density of observing in the stochastic process n of the continuous random variable X to take the values $x_1, x_2, \cdots, x_n$ respectively. Let us consider another stochastic process of the same continuous random variable X drawn according to a different probability distribution q, and let $q(x_1^n) = q(x_1, x_2, \cdots, x_n)$ be the joint probability distribution to observe the sequence $x_1, x_2, \cdots, x_n$ in the stochastic process drawn according to q. The KLD between the two probability distributions is expressed in the same way as the KLD between multivariate random variables (B.2),

$$D[p(x_1^n)\|q(x_1^n)] = \int dx_1^n \; p(x_1^n) \ln \frac{p(x_1^n)}{q(x_1^n)}, \tag{B.10}$$

In [Rol12] we introduced the following compact notation

$$D_n^X[p\|q] \equiv D[p(x_1^n)\|q(x_1^n)], \tag{B.11}$$

where the subscript n stands for the number of data sampled to distinguish between p and q and the superscript X denotes the variable that is distributed according to p and q. D_n^X is known as the $n-$time relative entropy between the distributions p and q of the random variable X. In case of stochastic process it is also interesting to define the KLD *rate* per data,

$$d^X[p\|q] = \lim_{n\to\infty} \frac{D_n^X[p\|q]}{n} \tag{B.12}$$

which saturates when n is of order of the memory of the stochastic process.

B.2 ANALYTICAL CALCULATION OF THE KLD BETWEEN HIDDEN MARKOV CHAINS USING REPLICA TRICK

B.2.1 *Replica trick*

Replica trick is a mathematical technique that was introduced to calculate free energies in spin glasses [VH79]. Consider a Hamiltonian that describes the interaction of N subsystems which are coupled by random strengths $\sigma_{i,j}$ that are distributed according to a specific probability distribution. One example is a spin glass with random coupling strengths between every

pair of spins, such as the Sherrington-Kirkpatrick model [She75] where coupling strengths are random and Gaussian distributed, $p(\sigma_{i,j}) \propto \exp(-\sigma_{i,j}^2/2)$. For such a system, it is interesting to calculate the free energy per subsystem. Let Z be the partition function of the $N-$spin system. The average free energy per site over all the possible values of the coupling, f, is given by

$$f = -k_B T \lim_{N\to\infty} \frac{1}{N} \langle \ln Z \rangle. \tag{B.13}$$

The average $\langle \ln Z \rangle$ over all possible configurations for these kind of models cannot be calculated analytically in general. However the following trick

Replica trick

$$\langle \ln Z \rangle = \lim_{n\to 0} \frac{d}{dn} \ln \langle Z^n \rangle, \tag{B.14}$$

simplifies the calculation of (B.13) since $\langle Z^n \rangle$ can be calculated explicitly in general. By replacing (B.14) into (B.13),

$$f = -k_B T \lim_{N\to\infty} \left(\lim_{n\to 0} \frac{d}{dn} \phi_N(n) \right), \tag{B.15}$$

where

$$\phi_N(n) = \frac{1}{N} \ln \langle Z^n \rangle, \tag{B.16}$$

can be considered as the free energy per site of a system formed by n replica of the original system. Because of this, the technique to replace the average $\langle \ln Z \rangle$ by the right hand side of Eq. (B.14) is often called *replica trick*.

B.2.2 *KLD of hidden Markov chains using replica trick*

The semi-analytical calculation of the KLD rate for a specific case of hidden Markov chains was discussed in Chapter 2.4.2. We now introduce a new technique to calculate the KLD d^X in Eq. (2.47) using the replica trick [Rol10, Rol12]. To this end, we first consider the expression of d^X in terms of Shannon and cross entropy rates, $d^X = h_r^X - h^X$. We define the matrix resulting from the multiplication of m transition matrices [defined in Eq. (2.54)] chosen according to x_1^m as

$$\mathbf{T}(x_1^m) = \prod_{i=1}^{m-1} \mathbf{T}(x_i, x_{i+1}), \tag{B.17}$$

Shannon entropy rate h^X can be rewritten by substituting (B.17) into (2.55),

$$h^X = -\lim_{m\to\infty} \frac{1}{m} \left\langle \ln \mathrm{Tr} \mathbf{T}(x_1^m) \right\rangle. \tag{B.18}$$

The analytical calculation of the average $\left\langle \ln \mathrm{Tr}\mathbf{T}(x_1^m) \right\rangle$ is cumbersome and it can only be done semi-analytically, as we explained in Chapter 2.4.2. However, we can express this average in terms of $\left\langle \mathrm{Tr}\mathbf{T}(x_1^m) \right\rangle$, which can be calculated analytically by using the replica trick [cf. Eq. (B.14)]

$$\langle \ln \mathrm{Tr}\mathbf{T}(x_1^m) \rangle = \lim_{\alpha \to 0} \frac{d}{d\alpha} \ln \langle [\mathrm{Tr}\mathbf{T}(x_1^m)]^\alpha \rangle. \tag{B.19}$$

Reference [DO96] shows how to apply this technique when $\mathbf{T}(x_1^m)$ is equal to a product of random matrices which are chosen following a Markovian process. In our case, an underlying Markovian process defined by the two random variables X and Y, defines the order of the matrices that are multiplied in $\mathbf{T}(x_1^m)$. We now apply the technique described in [DO96] to calculate h^X. If we define the *generalized Lyapunov exponent* of degree α [Cri93] as

$$L_\alpha^X = \lim_{m \to \infty} \frac{1}{m} \ln \left\langle [\mathrm{Tr}\mathbf{T}(x_1^m)]^\alpha \right\rangle, \tag{B.20}$$

which is analogous to the free energy per site described in (B.16). With this definition and taking into account the replica trick for this problem (B.19), Shannon entropy rate (B.18) is given by

$$h^X = - \lim_{\alpha \to 0} \frac{d}{d\alpha} L_\alpha^X. \tag{B.21}$$

Now we consider the following property: Given a matrix $\mathbf{A}$ and a positive integer α, $(\mathrm{Tr}\mathbf{A})^\alpha = \mathrm{Tr}(\mathbf{A}^{\otimes\alpha})$, where $\mathbf{A}^{\otimes\alpha} = \underbrace{\mathbf{A} \otimes \mathbf{A} \otimes \cdots \otimes \mathbf{A}}_{\alpha \text{ times}}$. Using this property, the average in (B.20) reads

$$\left\langle [\mathrm{Tr}\mathbf{T}(x_1^m)]^\alpha \right\rangle = \left\langle \mathrm{Tr}[\mathbf{T}(x_1^m)^{\otimes\alpha}] \right\rangle = \mathrm{Tr} \left\langle \mathbf{T}(x_1^m)^{\otimes\alpha} \right\rangle. \tag{B.22}$$

Since the tensor power of a product of matrices factorizes, $(\mathbf{ABC})^{\otimes\alpha} = \mathbf{A}^{\otimes\alpha}\mathbf{B}^{\otimes\alpha}\mathbf{C}^{\otimes\alpha}$, Eq. (B.22) can be rewritten,

$$\langle [\mathrm{Tr}\mathbf{T}(x_1^m)]^\alpha \rangle = \mathrm{Tr} \sum_{x_1^m, y_1^m} \prod_{i=1}^{m-1} \mathbf{T}(x_i, x_{i+1})_{y_i, y_{i+1}} \mathbf{T}(x_i, x_{i+1})^{\otimes\alpha}. \tag{B.23}$$

We now define a block matrix $\mathcal{T}(\alpha)$, where each block is a transition matrix $\mathbf{T}(x_1, x_2)^{\otimes\alpha+1}$. The matrix elements of $\mathcal{T}(\alpha)$ are therefore:

$$\mathcal{T}(\alpha)_{x_1, y_1, x_2, y_2} = [\mathbf{T}(x_1, x_2)^{\otimes\alpha+1}]_{y_1, y_2}. \tag{B.24}$$

Using (B.23) and (B.24) in (B.20), we see that L_α^X is dominated by the largest eigenvalue of $\mathcal{T}(\alpha)$ which we call $\tau(\alpha)$,

$$L_\alpha^X = \lim_{m \to \infty} \frac{1}{m} \ln \mathrm{Tr}[\mathcal{T}(\alpha)^{m-1}] = \ln \tau(\alpha), \tag{B.25}$$

yielding,

$$h^X = -\lim_{\alpha\to 0}\frac{d}{d\alpha}\ln\tau(\alpha). \tag{B.26}$$

The above limit cannot be calculated analytically because the tensor powers in $\mathcal{T}(\alpha)$ are only defined for integer values of α. Therefore we approximate the limit $\alpha \to 0$ by an estimation of the slope of L_α^X as a function of α close to $\alpha = 0$ [Cri93],

$$\hat{h}^X = 2L_1^X - \frac{L_2^X}{2} = 2\ln\tau(1) - \frac{\ln\tau(2)}{2}. \tag{B.27}$$

We obtain an equivalent result for h_r^X by replacing $\mathbf{T}(x_1^m)$ in (B.18) by the product of transition matrices but ordered according to the time-reversed series x_m^1, $\mathbf{T}(x_m^1)$. Defining the following matrix

$$\mathcal{T}_r(\alpha)_{x_1,y_1,x_2,y_2} = [\mathbf{T}(x_2,x_1)^\top \otimes \mathbf{T}(x_1,x_2)^{\otimes\alpha}]_{y_1,y_2}, \tag{B.28}$$

and $\tau_r(\alpha)$ being the largest eigenvalue of $\mathcal{T}_r(\alpha)$, we get

$$h_r^X = -\lim_{\alpha\to 0}\frac{d}{d\alpha}\ln\tau_r(\alpha). \tag{B.29}$$

In practice, we also need to approximate the limit $\alpha \to 0$ in the above expression using Eq. (B.27) but replacing τ by τ_r,

$$\hat{h}_r^X = 2\ln\tau_r(1) - \frac{\ln\tau_r(2)}{2}. \tag{B.30}$$

Finally, the estimation of d^X for this kind of series using replica trick, is obtained with the difference between Eqs. (B.30) and (B.27),

$$\hat{d}^X = \hat{h}_r^X - \hat{h}^X = 2\ln\frac{\tau_r(1)}{\tau(1)} + \frac{1}{2}\ln\frac{\tau(2)}{\tau_r(2)}. \tag{B.31}$$

C

APPENDIX OF CHAPTER 3

C.1 OUTGOING DEGREE DISTRIBUTION OF THE DHVG FOR AN UNCORRELATED RANDOM SERIES

We now proof Eq. (3.37) for real valued random uncorrelated series. We study the value of the *out* degree distribution, since the analysis for the *in* distribution is completely equivalent. Let x be an arbitrary data of the aforementioned series. The probability that the horizontal visibility of x is interrupted by a data x_r on its right is independent of the distribution that generates the series, $\rho(x)$,

$$\Phi_1 = \int_{-\infty}^{\infty}\int_{x}^{\infty}\rho(x)\rho(x_r)dx_r dx = \int_{-\infty}^{\infty}\rho(x)[1-F(x)]dx = \frac{1}{2}, \quad (C.1)$$

where $F(x) = \int_{-\infty}^{x}\rho(x')dx'$.

The probability $p_{out}(k)$ of the data x being capable of exactly seeing k data can be expressed as

$$p_{out}(k) = q_{out}(k)\Phi_1 = \frac{1}{2}q_{out}(k), \quad (C.2)$$

where $q_{out}(k)$ is the probability of x seeing at least k data. $q_{out}(k)$ can be recurrently calculated via

$$q_{out}(k) = q_{out}(k-1)(1-\Phi_1) = \frac{1}{2}q_{out}(k-1). \quad (C.3)$$

Since the first data at least sees the second data, $q_{out}(1) = 1$, the following expression is obtained

$$q_{out}(k) = \left(\frac{1}{2}\right)^{k-1}, \quad (C.4)$$

which together with equation (C.2) concludes the proof of Eq. (3.37),

$$p_{out}(k) = \left(\frac{1}{2}\right)^{k}. \quad (C.5)$$

We notice that an analogous derivation holds for the *in* case.

C.2 KLD VISIBILITY ESTIMATOR FOR A LOGISTIC MAP WTH $\mu = 4$.

In this section, we prove analytically that the time series generated by the logistic map (3.38) with $\mu = 4$ are irreversible at the level of

the *in* and *out* degree distributions of their associated DHVg, that is, $d_{vis} > 0$. We notice thatt the KLD between two distributions is zero if and only if the distributions are the same in the entire support. Therefore, if we want to prove that d_{vis} is strictly positive, it is sufficient to find that $P_{in}(k) \neq P_{out}(k)$ for some value of the degree k. Here we take advantage of this fact to provide a rather general recipe to prove that a chaotic system is irreversible.

Consider a time series x_1^n with a joint probability distribution $\rho(x_1, x_2, ..., x_n)$ and support (a, b), and denote x_{i-1}, x_i, x_{i+1} three (ordered) generic data of the series. By construction,

$$\begin{aligned} P_{out}(k=1) &= P(x_i \leqslant x_{i+1}) = \int_a^b dx_i \int_{x_i}^b dx_{i+1}\, \rho(x_i, x_{i+1}), \\ P_{in}(k=1) &= P(x_{i-1} > x_i) = \int_{x_i}^b dx_{i-1} \int_a^b dx_i \rho(x_{i-1}, x_i), \end{aligned} \tag{C.6}$$

The probability that $k_{out} = 1$ ($k_{in} = 1$) is actually the probability that the series increases (decreases) in one step. This probability is independent of time, because we consider stationary series. If the chaotic map is of the form $x_{i+1} = F(x_i)$, it is Markovian, and the preceding equations simplify due to Markov property (2.35):

$$\begin{aligned} P_{out}(k=1) &= \int_a^b dx_i \int_{x_t}^b dx_{i+1}\, \rho(x_i)\rho(x_{i+1}|x_i), \\ P_{in}(k=1) &= \int_a^b dx_i \int_{x_i}^b dx_{i-1}\, \rho(x_{i-1})\rho(x_i|x_{i-1}). \end{aligned} \tag{C.7}$$

For chaotic dynamical systems whose trajectories are in the attractor, there exists an invariant probability measure that characterizes the long-term fraction of time spent by the system in the various regions of the attractor. In the case of the Logistic map,

$$F(x_i) = \mu x_i(1 - x_i) \tag{C.8}$$

with parameter $\mu = 4$, the attractor is the whole interval $[0, 1]$ and the probability measure $\rho(x)$ corresponds to [Spro1]

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}. \tag{C.9}$$

Now, for a deterministic system, the transition probability is simply

$$\rho(x_{i+1}|x_i) = \delta(x_{i+1} - F(x_i)), \tag{C.10}$$

where $\delta(x)$ is the Dirac delta distribution. Equations (C.7) for the Logistic map with $\mu = 4$ and $x \in [0, 1]$ read

$$\begin{aligned} P_{out}(k=1) &= \int_0^1 dx_i \int_{x_i}^1 dx_{i+1}\, \rho(x_i)\delta(x_{i+1} - F(x_i)), \\ P_{in}(k=1) &= \int_0^1 dx_i \int_{x_i}^1 dx_{i-1}\, \rho(x_{i-1})\delta(x_i - F(x_{i-1})). \end{aligned} \tag{C.11}$$

Notice that, using the properties of the Dirac delta distribution, $\int_{x_i}^{1} \delta(x_{i+1} - F(x_i))dx_{i+1}$ is equal to one if and only if $F(x_i) \in [x_i, 1]$, what happens iff $0 < x_i < 3/4$, and it is zero otherwise. Therefore the only effect of this integral is to restrict the integration range of x_i to be $[0, 3/4]$. The first equation in (C.11) reduces to

$$P_{out}(k = 1) = \int_0^{3/4} dx_i \rho(x_i) = \frac{2}{3}. \tag{C.12}$$

On the other hand,

$$\int_{x_i}^{1} dx_{i-1} \rho(x_{i-1}) \delta(x_i - F(x_{i-1})) = \sum_{x_k^* | F(x_k^*) = x_i} \rho(x_k^*)/|F'(x_k^*)|, \tag{C.13}$$

that is, the sum over the roots x^* of the equation $F(x) = x$, iff $F(x_{i-1}) > x_0$. Since $x_{i-1} \in [x_i, 1]$ in the latter integral, the condition $F(x_{i-1}) > x_0$ is verified if $0 < x_i < 3/4$. In fact, if $0 < x_i < 3/4$ there is always a *single* value of $x_{i-1} \in [x_i, 1]$ such that $F(x_{i-1}) = x_i$, so the sum restricts to the adequate root. The particular value is $x_i^* = (1 + \sqrt{1 - x_i})/2$. Making use of these piecewise solutions and equation C.9, we finally have

$$P_{in}(k = 1) = \int_0^{3/4} dx_i \frac{\rho(x_i^*)}{4\sqrt{1 - x_i}} = \frac{1}{2} \int_0^{3/4} dx_i \rho(x_i) = \frac{1}{3}. \tag{C.14}$$

We conclude that $P_{out}(k) \neq P_{in}(k)$ for the Logistic map and hence the KLD measure based on degree distributions is positive.

Recall that $P_{out}(k = 1) = 2/3$ is the probability that the series exhibits a positive jump ($x_i > x_{i-1}$) once in the attractor. These positive jumps must be smaller in size than the negative jumps because, once in the attractor, $\langle x_i \rangle$ is constant. The irreversibility captured by the difference between $P_{out}(k = 1)$ and $P_{in}(k = 1)$ is then the asymmetry of the probability distribution of the slope $x_i - x_{i-1}$ of the original time series. The KLD of the degree distributions given by (3.34) clearly goes beyond this simple signature of irreversibility and can capture more complex and long-range traits.

D

APPENDIX OF CHAPTER 5

D.1 NUMERICAL INTEGRATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

We are interested in integrating numerically equations of the type

$$\dot{x}(t) = f(x(t);t) + g(x(t);t)\xi(t) \tag{D.1}$$

for any generic functions f and g that depend on x and could also depend explicitly on t. We focus on the one-dimensional case for simplicity. In the above equation, $\xi(t)$ is a Gaussian white noise of intensity σ (see Appendix A.1). We want to obtain a numerical scheme that approximates the solution of this kind of equations has the form

$$x(t) = x(0) + \int_0^t f(x(t');t')dt' + \int_0^t g(x(t');t')dW(t'). \tag{D.2}$$

We now show two simple numerical schemes that perform this task when (D.1) is taken in Ito or Stratonovich sense [Rüm82]. Interestingly, the two methods apply also to both Ito and Stratonovich case in case of additive noise.

D.1.1 *Euler's method*

Euler's method uses the simplest integration scheme, where the value of x at time $t + \Delta t$ —with Δt is smaller that any characteristic time of the system— is given by

$$x_{t+\Delta t} = x_t + f(x_t;t)\Delta t + g(x_t;t)\Delta W. \tag{D.3}$$

In the above equation, ΔW is the increment of the Wiener process in Δt, which is given by $\Delta W = \sigma\sqrt{\Delta t}\mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ is a random number generated by a Gaussian distribution of zero mean and unit variance. This numerical scheme converges to the Ito solution of (D.1) [Rüm82].

D.1.2 *Heun's method*

Heun's scheme is a refinement of Euler's method

$$x_{t+\Delta t} = x_t + \frac{1}{2}\left[f(x_t;t) + f(\tilde{x}_{t+\Delta t};t+\Delta t)\right]\Delta t + \frac{1}{2}\left[g(x_t;t) + g(\tilde{x}_{t+\Delta t};t+\Delta t)\right]\Delta W, \tag{D.4}$$

where $\tilde{x}_{t+\Delta t}$ is the predicted value of x at time $t + \Delta t$ using Euler's scheme,

$$\tilde{x}_{t+\Delta t} = x_t + f(x_t; t)\Delta t + g(x_t; t)\Delta W. \tag{D.5}$$

The numerical integration method given by Eqs. (D.4) and (D.5) converges to the Stratonovich solution of (D.1) as indicated in [Rüm82].

E

APPENDIX TO CHAPTER 7

E.1 POWER SPECTRUM DENSITY OF A BROWNIAN PARTICLE TRAPPED IN A HARMONIC TRAP

Let us consider an overdamped Brownian particle that moves in one dimension and that is trapped by a harmonic trap defined by the optical potential $V(x) = \frac{1}{2}\kappa x^2$. The position of the Brownian particle obeys the overdamped Langevin equation

$$\gamma \dot{x}(t) = -\kappa x(t) + \xi(t), \tag{E.1}$$

where $\xi(t)$ models the thermal noise, *i.e.* it is a Gaussian white noise with zero mean $\langle \xi(t) \rangle = 0$ and correlation

$$\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t - t'). \tag{E.2}$$

The Fourier transform of (E.1) is

$$i\omega \tilde{x}(\omega) = -\kappa \tilde{x}(\omega) + \tilde{\xi}(\omega) \tag{E.3}$$

where $\tilde{x}(\omega) = \int_{-\infty}^{\infty} dt e^{-i\omega t} x(t)$, ω being the angular frequency and $i = \sqrt{-1}$ is the imaginary unit. The PSD of the position of the Brownian particle is defined as

$$\mathrm{PSD}_x(\omega) = \frac{|\tilde{x}(\omega)|^2}{T_{msr}}, \tag{E.4}$$

where T_{msr} is the measurement time. From (E.3) and (E.2) we obtain

$$\mathrm{PSD}_x(\omega) = \frac{2k_B T\gamma}{\omega^2\gamma^2 + \kappa^2}. \tag{E.5}$$

If we define the *corner frequency* as $\omega_c = \kappa/\gamma$ one immediately gets

$$\mathrm{PSD}_x(\omega) = \frac{2k_B T}{\gamma} \frac{1}{\omega^2 + \omega_c^2}. \tag{E.6}$$

Equivalently using $f = 2\pi\omega$ and $f_c = 2\pi\omega_c$

$$\mathrm{PSD}_x(f) = \frac{k_B T}{2\pi^2\gamma} \frac{1}{f^2 + f_c^2}, \tag{E.7}$$

which proves Eq. (7.4).

E.2 CALCULATION OF THE EFFECTIVE KINETIC TEMPERATURE FROM THE POSITION HISTOGRAMS

We now prove the analytical expression (7.9) that relates T_{hist} with the properties of the additional random force applied in the experimental setup used in Chapter 7.

We consider a Brownian particle whose motion is described by the Langevin equation (7.2),

$$\gamma\dot{x}(t) = -\kappa x(t) + \xi(t) + \zeta(t), \tag{E.8}$$

where $\xi(t)$ and $\zeta(t)$ are Gaussian noises with zero average $\langle\xi(t)\rangle = \langle\zeta(t)\rangle = 0$ and their correlation functions are

$$\langle\xi(t)\xi(t')\rangle = 2k_B T\gamma\delta(t-t'), \tag{E.9}$$
$$\langle\zeta(t)\zeta(t')\rangle = \sigma^2\Gamma(t-t'). \tag{E.10}$$

The solution of Eq. (E.8) is:

$$x(t) = x(0)e^{-t/\tau_r} + \frac{e^{-t/\tau_r}}{\gamma}\int_0^t ds\; e^{s/\tau_r}[\xi(s)+\zeta(s)], \tag{E.11}$$

where $\tau_r = \gamma/\kappa$ is the relaxation time in the trap. If we multiply $x(t)$ by Eq. (E.8), and we do the average over different trajectories, we get

$$\frac{\gamma}{2}\frac{d\langle x^2\rangle}{dt} = -\kappa\langle x^2\rangle + \langle\xi x\rangle + \langle\zeta x\rangle, \tag{E.12}$$

where $\langle\xi x\rangle$ and $\langle\zeta x\rangle$ can be calculated using Novikov's theorem [Gol92, Iba00, Nov65a]:

$$\langle\zeta(t)x(t)\rangle = \int_0^t ds\langle\zeta(t)\zeta(s)\rangle\left\langle\frac{\delta x(t)}{\delta\zeta(s)}\right\rangle, \tag{E.13}$$

where $\frac{\delta x(t)}{\delta\zeta(s)}$ stands for functional the functional derivative of $x(t)$ with respect to $\zeta(s)$. Using Eq. (E.11), $\left\langle\frac{\delta x(t)}{\delta\zeta(s)}\right\rangle = \frac{1}{\gamma}e^{-(t-s)/\tau_r}$, therefore

$$\langle\zeta(t)x(t)\rangle = \frac{\sigma^2}{\gamma}\int_0^t ds\;\Gamma(t-s)e^{-(t-s)/\tau_r}, \tag{E.14}$$

whereas for the thermal noise

$$\langle\xi(t)x(t)\rangle = \int_0^t ds\langle\xi(t)\xi(s)\rangle\left\langle\frac{\delta x}{\delta\xi(s)}\right\rangle = k_B T. \tag{E.15}$$

Using Eqs. (E.14) and (E.15), Equation (E.12) can be rewritten as follows,

$$\frac{\gamma}{2}\frac{d\langle x^2\rangle}{dt} = -\kappa\langle x^2\rangle + k_B T + \frac{\sigma^2}{\gamma}\int_0^t dt'\;\Gamma(t')e^{-t'/\tau_r}. \tag{E.16}$$

In the steady state $t \to \infty$ and $d\langle x^2\rangle/dt = 0$. In this limit, the above equation yields (7.8)

$$\kappa\langle x^2\rangle_{ss} = k_B T + \frac{\sigma^2}{\gamma}\int_0^\infty dt'\, \Gamma(t')e^{-t'/\tau_r}. \tag{E.17}$$

By using equipartition theorem, we get T_{hist} as a function of the mean square displacement in the steady state,

$$T_{hist} = \frac{\kappa\langle x^2\rangle_{ss}}{k_B} = T + \frac{\sigma^2}{\gamma k_B}\int_0^\infty dt'\, \Gamma(t')e^{-t'/\tau_r}, \tag{E.18}$$

which proves Eq. (7.9).

We notice that $\Gamma(t)$ has a characteristic timescale given by τ_c, so it can be expressed as a function of t/τ_c, say $\Gamma(t) = \Gamma(t/\tau_c)$. The correlation of the noise decays in this timescale, which makes that the integral in the above equation can be expressed, by using the change of variable $s = t/\tau_r$,

$$\int_0^\infty dt'\, \Gamma(t'/\tau_c)e^{-t'/\tau_r} = \tau_r\int_0^\infty ds\, \Gamma\left(\frac{\tau_r}{\tau_c}s\right)e^{-s}. \tag{E.19}$$

If $\tau_c \ll \tau_r$, the exponential decays much slower with s than the correlation of the noise in the units above, therefore it can be approximated by $e^{-s} \simeq 1 - s$,

$$\int_0^\infty dt'\, \Gamma(t'/\tau_c)e^{-t'/\tau_r} \simeq \tau_r\int_0^\infty ds\, \Gamma\left(\frac{\tau_r}{\tau_c}s\right)(1-s) = \frac{1}{2} - \frac{\tau_c}{\tau_r}. \tag{E.20}$$

For the last equality, we have used that $\Gamma(t)$ is normalized, $\int_{-\infty}^{\infty} dt\Gamma(t/\tau_c) = 1$, that it is symmetric around $t = 0$ and the definition of the correlation time of the noise $\tau_c = \int_0^\infty dt\; t\; \Gamma(t/\tau_c)$. By using Eq. (E.20) into Eq. (E.18), we prove Eq. (7.10)

$$T_{hist} \simeq \frac{\sigma^2}{2\gamma k_B}\left(1 - \frac{2\tau_c}{\tau_r}\right). \tag{E.21}$$

E.3 CALCULATION OF THE AVERAGE WORK FOR A DRAGGED BROWNIAN PARTICLE IN A HARMONIC POTENTIAL

We now calculate the average of the work done on an overdamped Brownian particle trapped with a harmonic potential when the trap centre is moved at a constant velocity v for a period of time t. In this situation, the work done to move a trap that creates a moving quadratic potential $V(x(t), x_0(t)) = \frac{\kappa}{2}(x - x_0(t))^2$ is equal to

$$W = \int_0^t -\kappa(x - vt')v dt' = -\kappa v\int_0^t y(t')dt' \tag{E.22}$$

where we have used the definition of the work in Eq. (1.62) and $x_0(t) = v\ t$ as control parameter, and we have introduced the variable $y(t) = x(t) - vt$. $y(t)$ satisfies the following Langevin equation

$$\gamma\dot{y}(t) = -\kappa y(t) - \gamma v + \xi(t) + \zeta(t), \tag{E.23}$$

whose solution is

$$y(t) = y(0)e^{-t/\tau_r} + \frac{e^{-t/\tau_r}}{\gamma}\int_0^t ds\ e^{s/\tau_r}[-\gamma v + \xi(s) + \zeta(s)]. \tag{E.24}$$

The average over trajectories is given by

$$\langle y(t)\rangle = \langle y(0)\rangle e^{-t/\tau_r} - v\tau_r\left[1 - e^{-t/\tau_r}\right] = -v\tau_r\left[1 - e^{-t/\tau_r}\right], \tag{E.25}$$

where $\langle y(0)\rangle = \langle x(0)\rangle = 0$, since the bead is initially in equilibrium oscillating around the trap center located in $x = 0$. The average work reads

$$\langle W\rangle = -\kappa v\int_0^t \langle y(t')\rangle dt' = \gamma v^2 t + \frac{\gamma^2 v^2}{\kappa}\left[e^{-t/\tau_r} - 1\right], \tag{E.26}$$

which is the expression that we introduced in Eq. (7.12).

E.4 CALCULATION OF EFFECTIVE CROOKS TEMPERATURE FOR A DRAGGED BROWNIAN PARTICLE

For a process in which a physical system that starts in equilibrium it is driven out of equilibrium such that the probability distribution of the work is Gaussian, Crooks temperature is equal to Eq. (7.14)

$$T_C = \frac{\sigma_W^2}{2k_B\langle W\rangle}, \tag{E.27}$$

being $\langle W\rangle$ given by Eq. (7.12). We are now interested in calculating the variance of the work done in the nonequilibrium process consisting in moving the trap at a constant velocity v during a time t, which we call σ_W^2. We first notice that the work is defined in terms of the variable $y(t) = x(t) - vt$ as shown in Eq. (E.22). If we introduce the random variables

$$q_\xi(t') = \int_0^{t'} ds\ e^{-(t'-s)/\tau_r}\xi(s), \quad q_\zeta(t') = \int_0^{t'} ds\ e^{-(t'-s)/\tau_r}\zeta(s), \tag{E.28}$$

$y(t)$ can be rewritten as

$$y(t) = y(0)e^{-t/\tau_r} - v\tau_r\left[1 - e^{-t/\tau_r}\right] + \frac{1}{\gamma}\left[q_\xi(t) + q_\eta(t)\right]. \tag{E.29}$$

By replacing the above formula in Eq. (E.22), we get

$$W = -\gamma\, v\, y(0)\left[1 - e^{-t/\tau_r}\right] + \langle W\rangle - \frac{\kappa v}{\gamma}\int_0^t dt'[q_\xi(t') + q_\zeta(t')]. \tag{E.30}$$

Therefore, the work is a linear combination of random variables that are independent each other. This implies that the variance of the work can be expressed as the following sum

$$\sigma_W^2 = \gamma^2 v^2\left[1 - e^{-t/\tau_r}\right]^2 \sigma_{x(0)}^2 + \frac{\kappa^2 v^2}{\gamma^2}\sigma_{q_\xi}^2 + \frac{\kappa^2 v^2}{\gamma^2}\sigma_{q_\zeta}^2. \tag{E.31}$$

The first term concerns the variance of the position at $t = 0$ in equilibrium, $\sigma_{x(0)}^2$, which is known (E.18).

$$\sigma_{x(0)}^2 = \frac{k_B T_{hist}}{\kappa}, \tag{E.32}$$

The calculation of $\sigma_{q_\xi}^2$ and $\sigma_{q_\zeta}^2$ is not straightforward. The first can be calculated analytically whereas the second only for particular correlation functions of the external noise. We now show how a closed expression for the term concerning the thermal noise, $\sigma_{q_\xi}^2$, can be derived. This variance is defined as

$$\sigma_{q_\xi}^2 = \int_0^t dt_1 \int_0^t dt_2 \langle q_\xi(t_1) q_\xi(t_2)\rangle \tag{E.33}$$

where

$$\langle q_\xi(t_1) q_\xi(t_2)\rangle = \sigma^2 \int_0^{t_1} dt' \int_0^{t_2} dt'' e^{-(t_1 - t')/\tau_r} e^{-(t_2 - t'')/\tau_r} \Gamma(t' - t''). \tag{E.34}$$

For thermal Gaussian white noise, $\sigma^2 = 2\gamma k_B T$ and $\Gamma(t' - t'') = \delta(t' - t'')$, which yields

$$\langle q_\xi(t_1) q_\xi(t_2)\rangle = k_B T \frac{\gamma^2}{\kappa}\left[e^{-|t_1 - t_2|/\tau_r} - e^{-(t_1 + t_2)/\tau_r}\right]. \tag{E.35}$$

By replacing the above result into Eq. (E.33), and integrating, we obtain

$$\sigma_{q_\xi}^2 = 2k_B T \frac{\gamma^3}{\kappa^2} t + k_B T \frac{\gamma^4}{\kappa^3}\left[4e^{-t/\tau_r} - e^{-2t/\tau_r} - 3\right]. \tag{E.36}$$

After some algebra, T_C can be expressed as a function of $T_e = \sigma^2/2k_B\gamma$ for any correlation function $\Gamma(t)$ as follows:

$$T_c = T + \frac{\mathcal{L}\{\Gamma(t)\}(1/\tau_r) + \frac{1}{\tau_r^3}\int_0^t dt_1 \int_0^t dt_2 \int_0^{t_1} dt' \int_0^{t_2} dt'' e^{-(t_1 - t')/\tau_r} e^{-(t_2 - t'')/\tau_r} \Gamma(t' - t'')}{\frac{t}{\tau_r} + e^{-t/\tau_r} - 1} T_e, \tag{E.37}$$

where $\mathcal{L}\{\Gamma(t)\}(1/\tau_r)$ is the Laplace transform of $\Gamma(t)$ evaluated at $s = 1/\tau_r$,

$$\mathcal{L}\{\Gamma(t)\}(1/\tau_r) \equiv \int_0^\infty dt'\, \Gamma(t') e^{-t'/\tau_r}. \tag{E.38}$$

For the correlation function that fits the experimental data (7.11), Eq. (E.37) can only be calculated numerically. In the limit in which the external noise is white, *i.e.* $\sigma^2 = 2\gamma k_B T_e$ and $\Gamma(t) = \delta(t)$, we obtain an analogous result to (E.36) but replacing T by T_e. Only in this case $T_C = T_{hist} = T_{PSD} = T + T_e$ for any value of the driving time τ.

BIBLIOGRAPHY

[Ajd92] A. Ajdari and J. Prost: *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre* **315**, 1635–1639 (1992).

[Ale12] A. Alemany, A. Mossa, I. Junier and F. Ritort: *Nature Physics* **8**, 688–694 (2012).

[And07] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud and A. Petrosyan: *Physical review letters* **98**, 150601 (2007).

[And08a] D. Andrieux and P. Gaspard: *Proceedings of the National Academy of Sciences* **105**, 9516–9521 (2008).

[And08b] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud and A. Petrosyan: *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P01002 (2008).

[Ash77] A. Ashkin and J. Dziedzic: *Applied Physics Letters* **30**, 202–204 (1977).

[Ash79] N. W. Ashcroft and N. D. Mermin: *There is no corresponding record for this reference* (1979).

[Ash86] A. Ashkin, J. Dziedzic, J. Bjorkholm and S. Chu: *Optics letters* **11**, 288–290 (1986).

[Bal75] R. Balescu: *NASA STI/Recon Technical Report A* **76**, 32809 (1975).

[Bat00] G. K. Batchelor: *An introduction to fluid dynamics* (Cambridge university press, 2000).

[Bay72] R. Bayer and E. McCreight: *Acta Informatica* **1**, 173–189 (1972).

[Ben02] D. Benedetto, E. Caglioti and V. Loreto: *Physical Review Letters* **88**, 048702 (2002).

[Bér12] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider and E. Lutz: *Nature* **483**, 187–189 (2012).

[Beu12] F. Beunis, F. Strubbe, K. Neyts and D. Petrov: *Physical review letters* **108**, 016101 (2012).

[Bli11] V. Blickle and C. Bechinger: *Nature Physics* (2011).

[Box11] G. E. Box, G. M. Jenkins and G. C. Reinsel: *Time series analysis: forecasting and control*, volume 734 (Wiley, 2011).

[Bro28] R. Brown: *The Philosophical Magazine, or Annals of Chemistry, Mathematics, Astronomy, Natural History and General Science* **4**, 161–173 (1828).

[Bus01] C. Bustamante, D. Keller and G. Oster: *Accounts of chemical research* **34**, 412–420 (2001).

[Bus03] C. Bustamante, Z. Bryant, S. Smith et al.: *Nature* pages 423–426 (2003).

[Bus05] C. Bustamante, J. Liphardt and F. Ritort Farran: *Physics Today, 2005, vol. 58, num. 7, p. 43-48* (2005).

[Cai06] H. Cai, S. R. Kulkarni and S. Verdú: *Information Theory, IEEE Transactions on* **52**, 3456–3475 (2006).

[Cal85] H. B. Callen: *Thermodynamics and an Introduction to Thermostatistics, 2nd Edition, by Herbert B. Callen, pp. 512. ISBN 0-471-86256-8. Wiley-VCH, August 1985.* **1** (1985).

[Cec05] C. Cecconi, E. A. Shank, C. Bustamante and S. Marqusee: *Science* **309**, 2057–2060 (2005).

[Cil98] S. Ciliberto and C. Laroche: *Le Journal de Physique IV* **8** (1998).

[Cil04] S. Ciliberto, N. Garnier, S. Hernandez, C. Lacpatia, J.-F. Pinton and G. Ruiz Chavarria: *Physica A: Statistical Mechanics and its Applications* **340**, 240–250 (2004).

[Cil10] S. Ciliberto, S. Joubaud and A. Petrosyan: *Journal of Statistical Mechanics: Theory and Experiment* **2010**, P12003 (2010).

[Cle06] B. Cleuren, C. Van den Broeck and R. Kawai: *Physical review letters* **96**, 050601 (2006).

[Coc] http://nl.bu.edu/wp-content/uploads/2011/10/i10-85-cochlea22.jpg.

[Cof04] W. Coffey, Y. P. Kalmykov and J. T. Waldron: *The Langevin equation: with applications to stochastic problems in physics, chemistry, and electrical engineering*, volume 14 (World Scientific Publishing Company Incorporated, 2004).

[Col05] D. Collin, F. Ritort, C. Jarzynski, S. Smith, I. Tinoco and C. Bustamante: *Nature* **437**, 231–234 (2005).

[Cos05] M. Costa, A. L. Goldberger and C.-K. Peng: *Physical review letters* **95**, 198102 (2005).

[Cos08] M. D. Costa, C.-K. Peng and A. L. Goldberger: *Cardiovascular Engineering* **8**, 88–93 (2008).

[Cou05] D. P. Coutinho and M. A. Figueiredo: In *Pattern Recognition and Image Analysis*, pages 355–362 (Springer, 2005).

[Cou10] D. P. Coutinho, A. L. Fred and M. A. Figueiredo: In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3858–3861 (IEEE, 2010).

[Cov06] T. M. Cover and J. A. Thomas: *Elements of information theory* (Wiley-interscience, 2006).

[Cri93] A. Crisanti, G. Paladin and A. Vulpiani: *Products of Random Matrices in Statistical Physics (Springer Series in Solid State Sciences)* (Springer, 1993). ISBN 0387565752.

[Cro98] G. E. Crooks: *Journal of Statistical Physics* **90**, 1481–1487 (1998).

[Cro99] G. Crooks: *Physical Review E* **60**, 2721 (1999).

[DG11] S. R. De Groot and P. Mazur: *Non-equilibrium thermodynamics* (Dover publications, 2011).

[Die04] H. Dietz and M. Rief: *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16192–16197 (2004).

[Din07] L. Dinis, E. González, J. V. Anguita, J. Parrondo and J. Vicent: *New Journal of Physics* **9**, 366 (2007).

[DO96] M. De Oliveira and A. Petri: *Physical Review E* **53**, 2960 (1996).

[Dor10] S. N. Dorogovtsev: *Lectures on complex networks* (Oxford University Press London, 2010).

[Duc91] W. Ducker, T. Senden and R. Pashley: *Nature* **353**, 239–241 (1991).

[Dun61] E. Dunham and I. Glynn: *The Journal of physiology* **156**, 274–293 (1961).

[Ear] http://upload.wikimedia.org/wikipedia/commons/d/d2/Anatomy_of_the_Human_Ear.svg.

[Ein05] A. Einstein: *Ann. Phys., Lpz* **17**, 549 (1905).

[Els09] J. Elsner, T. Jagger and E. Fogarty: *Geophysical Research Letters* **36** (2009).

[Erd59] P. Erdős and A. Rényi: *Publicationes Mathematicae Debrecen* **6**, 290–297 (1959).

[Erd61] P. Erdős and A. Renyi: *Acta Mathematica Hungarica* **12**, 261–267 (1961).

[Eva93] D. J. Evans, E. Cohen and G. Morriss: *Physical Review Letters* **71**, 2401–2404 (1993).

[Eva96] D. J. Evans and D. J. Searles: *Physical Review E* **53**, 5808 (1996).

[Fei04] K. Feitosa and N. Menon: *Physical review letters* **92**, 164301 (2004).

[Fer56] E. Fermi: *Thermodynamics, 1936* (1956).

[Fey65] R. P. Feynman, R. B. Leighton, M. Sands and E. Hafner: *American Journal of Physics* **33**, 750 (1965).

[Gal95] G. Gallavotti and E. Cohen: *Physical Review Letters* **74**, 2694–2697 (1995).

[Gar05] N. Garnier and S. Ciliberto: *Physical Review E* **71**, 060101 (2005).

[Gas04] P. Gaspard: *Journal of statistical physics* **117**, 599–615 (2004).

[Gie12] J. Gieseler, B. Deutsch, R. Quidant and L. Novotny: *Physical Review Letters* **109**, 103603 (2012).

[GM08a] A. Gomez-Marin, J. Parrondo and C. Van den Broeck: *Physical Review E* **78**, 011107 (2008).

[GM08b] A. Gomez-Marin, J. Parrondo and C. Van den Broeck: *EPL-Europhysics Letters* **82**, 50002–54000 (2008).

[Gol92] N. Goldenfeld: (1992).

[Gos02] C. Gosse and V. Croquette: *Biophysical journal* **82**, 3314–3329 (2002).

[Gra03] P. Grassberger: *arXiv preprint physics/0307138* (2003).

[Gre95] W. Greiner, L. Neise, H. Stöcker and D. Rischke: *Thermodynamics and statistical mechanics*, volume 1 (Springer-Verlag New York, 1995).

[Gri92] G. Grimmett and D. Stirzaker: *Probability and random processes*, volume 2 (Clarendon press Oxford, 1992).

[GS10] J. R. Gomez-Solano, L. Bellon, A. Petrosyan and S. Ciliberto: *EPL (Europhysics Letters)* **89**, 60003 (2010).

[Gut11] G. Gutin, T. Mansour and S. Severini: *Physica A: Statistical Mechanics and its Applications* **390**, 2421–2428 (2011).

[Hän09] P. Hänggi and F. Marchesoni: *Reviews of Modern Physics* **81**, 387 (2009).

[Has70] W. K. Hastings: *Biometrika* **57**, 97–109 (1970).

[HbW] http://michaeldmann.net/pix_4b/hair_cell.gif.

[Hol06] T. Holliday, A. Goldsmith and P. Glynn: *Information Theory, IEEE Transactions on* **52**, 3509–3532 (2006).

[Hor09] J. Horowitz and C. Jarzynski: *Physical Review E* **79**, 21106 (2009).

[How87] J. Howard and A. Hudspeth: *Proceedings of the National Academy of Sciences* **84**, 3064–3068 (1987).

[How01] J. Howard: *Mechanics of motor proteins and the cytoskeleton* (Sinauer Associates Sunderland, MA, 2001).

[Hua87] K. Huang: Statistical mechanics, 18.3 (1987).

[Hud08] A. Hudspeth: *Neuron* **59**, 530–545 (2008).

[Hux69] H. Huxley: *Science* **164**, 1356–1366 (1969).

[Iba00] M. Ibañes, J. García-Ojalvo, R. Toral and J. M. Sancho: In *Stochastic Processes in Physics, Chemistry, and Biology*, pages 247–256 (Springer, 2000).

[Ion] http://nestbio.blogspot.com.es/2010/11/passive-and-active-transport.html.

[Jac08] P. Jacquet, G. Seroussi and W. Szpankowski: *Theoretical computer science* **395**, 203–219 (2008).

[Jah11] M. Jahnel, M. Behrndt, A. Jannasch, E. Schffer and S. W. Grill: *Optics letters* **36**, 1260 (2011).

[Jar97] C. Jarzynski: *Physical Review Letters* **78**, 2690 (1997).

[Jar99] C. Jarzynski and O. Mazonka: *Physical Review E* **59**, 6448 (1999).

[Jar08] C. Jarzynski: *The European Physical Journal B-Condensed Matter and Complex Systems* **64**, 331–340 (2008).

[Jar11] C. Jarzynski: *Annu. Rev. Condens. Matter Phys.* **2**, 329–351 (2011).

[Jop08] P. Jop, A. Petrosyan and S. Ciliberto: *EPL (Europhysics Letters)* **81**, 50005 (2008).

[Jül97a] F. Jülicher, A. Ajdari and J. Prost: *Reviews of Modern Physics* **69**, 1269 (1997).

[Jül97b] F. Jülicher, A. Ajdari and J. Prost: *Reviews of Modern Physics* **69**, 1269 (1997).

[Kaw07] R. Kawai, J. Parrondo and C. V. den Broeck: *Physical review letters* **98**, 80602 (2007).

[Kel99] T. R. Kelly, H. De Silva and R. A. Silva: *Nature* **401**, 150–152 (1999).

[Kin] http://www.stanford.edu/group/blocklab/ScienceLimping/StructureCartoon.jpg.

[Kin69] J. Kingman: *Journal of Applied Probability* pages 1–18 (1969).

[Kon98] D. Kondepudi and I. Prigogine: *From Heat Engines to Dissipative Structures* (John Wiley & Son, 1998).

[Kos93] E. J. Kostelich and T. Schreiber: *Physical Review E* **48**, 1752 (1993).

[Kra40] H. A. Kramers: *Physica* **7**, 284–304 (1940).

[Kra04] A. Kraskov, H. Stögbauer and P. Grassberger: *Physical Review E* **69**, 066138 (2004).

[Kub66] R. Kubo: *Reports on Progress in Physics* **29**, 255 (1966).

[Kul51] S. Kullback and R. A. Leibler: *The Annals of Mathematical Statistics* **22**, 79–86 (1951).

[Kur99] J. Kurchan: *Journal of Physics A: Mathematical and General* **31**, 3719 (1999).

[Lac08] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. C. Nuño: *Proceedings of the National Academy of Sciences* **105**, 4972–4975 (2008).

[Lac09a] L. Lacasa, B. Luque, J. Luque and J. C. Nuno: *EPL (Europhysics Letters)* **86**, 30001 (2009).

[Lac09b] D. Lacoste and K. Mallick: *Physical review. E, Statistical, nonlinear, and soft matter physics* **80**, 021923 (2009).

[Lac10] L. Lacasa and R. Toral: *Physical Review E* **82**, 036120 (2010).

[Lac12] L. Lacasa, A. Nuñez, É. Roldán, J. M. Parrondo and B. Luque: *The European Physical Journal B-Condensed Matter and Complex Systems* **85**, 1–11 (2012).

[Lan08] P. Langevin: *On the theory of Brownian motion* (1908).

[Lan61] R. Landauer: *IBM journal of research and development* **5**, 183–191 (1961).

[Leb99] J. L. Lebowitz and H. Spohn: *Journal of Statistical Physics* **95**, 333–365 (1999).

[Lef92] H. S. Leff, A. F. Rex and D. L. Hogenboom: *American Journal of Physics* **60**, 282 (1992).

[Li11] T. Li, S. Kheifets and M. G. Raizen: *Nature Physics* **7**, 527–530 (2011).

[Lin98] H. Linke, W. Sheng, A. Löfgren, H. Xu, P. Omling and P. Lindelof: *EPL (Europhysics Letters)* **44**, 341 (1998).

[Lin06] H. Linke, B. Alemán, L. Melling, M. Taormina, M. Francis, C. Dow-Hygelund, V. Narayanan, R. Taylor and A. Stout: *Physical review letters* **96**, 154502 (2006).

[Lip02] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr and C. Bustamante: *Science* **296**, 1832–1835 (2002).

[Liu10] C. Liu, W.-X. Zhou and W.-K. Yuan: *Physica A: Statistical Mechanics and its Applications* **389**, 2675–2681 (2010).

[Lóp12] D. J. López, M. Egido-Gabas, I. López-Montero, J. V. Busto, J. Casas, M. Garnier, F. Monroy, B. Larijani, F. M. Goñi and A. Alonso: *Biophysical Journal* **102**, 2077 (2012).

[Luq09] B. Luque, L. Lacasa, F. Ballesteros and J. Luque: *Physical Review E* **80**, 046103 (2009).

[Luq11] B. Luque, L. Lacasa, F. J. Ballesteros and A. Robledo: *PloS one* **6**, e22411 (2011).

[Mao06] H. Ma, D. J. Proctor, E. Kierzek, R. Kierzek, P. C. Bevilacqua and M. Gruebele: *Journal of the American Chemical Society* **128**, 1523–1530 (2006).

[Mae99] C. Maes: *Journal of Statistical Physics* **95**, 367–392 (1999).

[Mae03] C. Maes: *Séminaire Poincaré* **2**, 29–62 (2003).

[Man08] G. A. Manley, R. R. Fay and A. N. Popper: *Active processes and otoacoustic emissions*, volume 30 (Springer, 2008).

[Mao05] H. Mao, J. Ricardo Arias-Gonzalez, S. B. Smith, I. Tinoco Jr and C. Bustamante: *Biophysical journal* **89**, 1308–1316 (2005).

[Mar99] P. Martin and A. Hudspeth: *Proceedings of the National Academy of Sciences* **96**, 14306–14311 (1999).

[Mar01a] P. Martin and A. Hudspeth: *Proceedings of the National Academy of Sciences* **98**, 14386–14391 (2001).

[Mar01b] P. Martin, A. Hudspeth and F. Jülicher: *Proceedings of the National Academy of Sciences* **98**, 14380–14385 (2001).

[Mar03] P. Martin, D. Bozovic, Y. Choe and A. Hudspeth: *The Journal of neuroscience* **23**, 4533–4548 (2003).

[Mar08] P. Maragakis, M. Spichty and M. Karplus: *The Journal of Physical Chemistry B* **112**, 6168–6174 (2008).

[Mar12a] I. A. Martínez and D. Petrov: *Applied Optics* **51**, 5973–5977 (2012).

[Mar12b] I. A. Martínez and D. Petrov: *Applied Optics* **51**, 5522–5526 (2012).

[mat] MATLAB R2011b documentation (curve fitting toolbox). http://www.mathworks.es/help/toolbox/curvefit/.

[McC99] L. I. McCann, M. Dykman and B. Golding: *Nature* **402**, 785–787 (1999).

[Mes11] P. Mestres and D. Petrov: *European Biophysics Journal* **40**, 1081–1085 (2011).

[Mie08] G. Mie: *Annalen der Physik* **330**, 377–445 (1908).

[Nad04] B. Nadrowski, P. Martin and F. Jülicher: *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12195–12200 (2004).

[New03] M. E. Newman: *SIAM review* **45**, 167–256 (2003).

[Noj97] H. Noji, R. Yasuda, M. Yoshida and K. Kinosita: *Nature* **386**, 299–302 (1997).

[Nov65a] E. Novikov: *Sov. Phys. JETP* **20**, 1290–1294 (1965).

[Nov65b] E. A. Novikov: *Soviet Physics-JETP* **20**, 1290–1294 (1965).

[Onu97] J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes: *Annual Review of Physical Chemistry* **48**, 545–600 (1997).

[Par96] J. M. Parrondo and P. Español: *American Journal of Physics* **64**, 1125–1129 (1996).

[Par01] J. M. Parrondo: *Chaos: An Interdisciplinary Journal of Nonlinear Science* **11**, 725–733 (2001).

[Par02] J. Parrondo and B. J. de Cisneros: *Applied Physics A: Materials Science & Processing* **75**, 179–191 (2002).

[Par03] S. Park, F. Khalili-Araghi, E. Tajkhorshid and K. Schulten: *The Journal of chemical physics* **119**, 3559 (2003).

[Par04] J. Parrondo and L. Dínis: *Contemporary Physics* **45**, 147–157 (2004).

[Par09] J. M. Parrondo, C. Van den Broeck and R. Kawai: *New Journal of Physics* **11**, 073008 (2009).

[Pat88] R. Pathria: *Statistical Mechanics: International Series in Natural Philosophy* **45** (1988).

[PC08] F. Pérez-Cruz: In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 1666–1670 (IEEE, 2008).

[Pet13] D. Petrov: Teaching master in photonics "advanced experimental techniques in biology" (2013). URL http://users.icfo.es/Dmitri.Petrov/Teaching/lectures.htm.

[Pha] http://biomedicalcomputationreview.org/content/viral-closeup.

[Por07] A. Porporato, J. Rigby and E. Daly: *Physical review letters* **98**, 094101 (2007).

[Rab89] L. R. Rabiner: *Proceedings of the IEEE* **77**, 257–286 (1989).

[Rac04] Z. Rached, F. Alajaji and L. L. Campbell: *Information Theory, IEEE Transactions on* **50**, 917–921 (2004).

[Rad09] R. Radünz, D. Rings, K. Kroy and F. Cichos: *The Journal of Physical Chemistry A* **113**, 1674–1677 (2009).

[Rei02] P. Reimann: *Physics Reports* **361**, 57–265 (2002).

[Ric08] A. C. Richardson, S. Reihani and L. B. Oddershede: *Optics express* **16**, 15709–15717 (2008).

[Rol10] É. Roldán and J. M. Parrondo: *Physical review letters* **105**, 150607 (2010).

[Rol12] É. Roldán and J. M. Parrondo: *Physical Review E* **85**, 031129 (2012).

[Rui11] P. Ruijgrok, N. Verhart, P. Zijlstra, A. Tchebotareva and M. Orrit: *Physical Review Letters* **107**, 037401 (2011).

[Rüm82] W. Rümelin: *SIAM Journal on Numerical Analysis* pages 604–613 (1982).

[Sag08] T. Sagawa and M. Ueda: *Physical review letters* **100**, 080403 (2008).

[Sag09] T. Sagawa and M. Ueda: *Physical review letters* **102**, 250602 (2009).

[Sch96] T. Schürmann and P. Grassberger: *Chaos: An Interdisciplinary Journal of Nonlinear Science* **6**, 414–427 (1996).

[Sch05] S. Schuler, T. Speck, C. Tietz, J. Wrachtrup and U. Seifert: *Physical review letters* **94**, 180602 (2005).

[Sea99] D. J. Searles and D. J. Evans: *Physical Review E* **60**, 159 (1999).

[Sei05] U. Seifert: *Physical review letters* **95**, 40602 (2005).

[Sei08] U. Seifert: *The European Physical Journal B-Condensed Matter and Complex Systems* **64**, 423–431 (2008).

[Sei12] U. Seifert: *arXiv preprint arXiv:1205.4176* (2012).

[Sek98] K. Sekimoto: *Progress of Theoretical Physics Supplement* **130**, 17–27 (1998).

[Sek10] K. Sekimoto: *Stochastic energetics*, volume 799 (Springer, 2010).

[She75] D. Sherrington and S. Kirkpatrick: *Physical review letters* **35**, 1792 (1975).

[Sim92] A. Simon and A. Libchaber: *Physical review letters* **68**, 3375–3378 (1992).

[Smi01] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson and C. Bustamante: *Nature* **413**, 748–752 (2001).

[Smo17] M. v. Smoluchowski: *Z. phys. Chem* **92**, 9 (1917).

[Spr01] J. Sprott and G. Rowlands: *International Journal of Bifurcation and Chaos* **11**, 1865–1880 (2001).

[Str01] S. Strogatz: (2001).

[Svo93] K. Svoboda, C. F. Schmidt, B. J. Schnapp, S. M. Block et al.: *Nature* **365**, 721–727 (1993).

[Svo94a] K. Svoboda and S. M. Block: *Annual review of biophysics and biomolecular structure* **23**, 247–285 (1994).

[Svo94b] K. Svoboda and S. M. Block: *Cell* **77**, 773–784 (1994).

[Szi29] L. Szilard: *Zeitschrift für Physik A Hadrons and Nuclei* **53**, 840–856 (1929).

[Tor77] G. M. Torrie and J. P. Valleau: *Journal of Computational Physics* **23**, 187–199 (1977).

[Toy10] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki and M. Sano: *Nature Physics* **6**, 988–992 (2010).

[Ved02] V. Vedral: *Reviews of Modern Physics* **74**, 197 (2002).

[VH79] J. Van Hemmen and R. Palmer: *Journal of Physics A: Mathematical and General* **12**, 563 (1979).

[Vil03] J. Villegas, S. Savel'ev, F. Nori, E. Gonzalez, J. Anguita, R. Garcia and J. Vicent: *Science* **302**, 1188–1191 (2003).

[Vis96] K. Visscher, S. P. Gross and S. M. Block: *Selected Topics in Quantum Electronics, IEEE Journal of* **2**, 1066–1076 (1996).

[Vis99] K. Visscher, M. Schnitzer and S. Block: *Nature* **400**, 184–189 (1999).

[VK92] N. G. Van Kampen: *Stochastic processes in physics and chemistry*, volume 1 (North holland, 1992).

[Vol06] G. Volpe and D. Petrov: *Physical review letters* **97**, 210603 (2006).

[Vol08] G. Volpe, S. Perrone, J. M. Rubi and D. Petrov: *Physical Review E* **77**, 051107 (2008).

[VZ03a] R. Van Zon and E. Cohen: *Physical Review Letters* **91**, 110601 (2003).

[VZ03b] R. Van Zon and E. Cohen: *Physical Review E* **67**, 046102 (2003).

[Wan02] G. Wang, E. M. Sevick, E. Mittag, D. J. Searles and D. J. Evans: *Physical Review Letters* **89**, 50601 (2002).

[Wan05a] G. Wang, J. Reid, D. Carberry, D. Williams, E. M. Sevick and D. J. Evans: *Physical Review E* **71**, 046142 (2005).

[Wan05b] Q. Wang, S. R. Kulkarni and S. Verdú: *Information Theory, IEEE Transactions on* **51**, 3064–3074 (2005).

[Wan09] Q. Wang, S. R. Kulkarni and S. Verdú: *IEEE Transactions on Information Theory* **55**, 2392–2405 (2009).

[Wat98] D. J. Watts and S. H. Strogatz: *nature* **393**, 440–442 (1998).

[Wei75] G. Weiss: *Journal of Applied Probability* pages 831–836 (1975).

[Yan09] Y. Yang, J. Wang, H. Yang and J. Mang: *Physica A: Statistical Mechanics and its Applications* **388**, 4431–4437 (2009).

[Zem68] M. W. Zemansky and R. Dittman: *Heat and thermodynamics: an intermediate textbook* (McGraw-Hill, 1968).

[Ziv78] J. Ziv and A. Lempel: *Information Theory, IEEE Transactions on* **24**, 530–536 (1978).

[Ziv93] J. Ziv and N. Merhav: *Information Theory, IEEE Transactions on* **39**, 1270–1279 (1993).

[Zur03] W. H. Zurek: *arXiv preprint quant-ph/0301076* (2003).

[Zwa73] R. Zwanzig: *Journal of Statistical Physics* **9**, 215–220 (1973).

[Zwa01] R. Zwanzig: *Nonequilibrium statistical mechanics* (Oxford University Press, USA, 2001).